

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:03 : Search time 43.6471 Seconds
(without alignments)
285.020 Million cell updates/sec

Title: US-10-032-658-11
Perfect score: 664
Sequence: 1 MARFCTGFSKKMLVIAVIWM.....DSNCKYKATVACNTSGCPGH 112

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: A_Geneseq_032802:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 664 | 100.0 | 112 | 20 | AAV09278 |
| 2 | 603 | 90.8 | 124 | 20 | AAV09281 |
| 3 | 599 | 90.2 | 112 | 20 | AAV09279 |
| 4 | 599 | 90.2 | 124 | 20 | AAV09282 |
| 5 | 487 | 73.3 | 148 | 20 | AAV09280 |
| 6 | 377.5 | 56.9 | 108 | 18 | AAW07693 |
| 7 | 361 | 54.4 | 109 | 18 | AAW07694 |
| 8 | 156 | 23.5 | 1679 | 22 | AAW07343 |
| 9 | 139 | 20.9 | 233 | 21 | AAV74791 |
| 10 | 124 | 18.7 | 24 | 20 | AAV09276 |
| 11 | 123.5 | 18.6 | 524 | 22 | AAW07370 |

| | | | | | |
|----|-------|------|------|----|----------|
| 12 | 119 | 17.9 | 250 | 21 | AAV82335 |
| 13 | 118.5 | 17.8 | 2189 | 21 | AAV05222 |
| 14 | 116 | 17.5 | 1700 | 21 | AAV18144 |
| 15 | 111 | 16.7 | 225 | 21 | AAV36084 |
| 16 | 110.5 | 16.6 | 233 | 20 | AAV29053 |
| 17 | 110.5 | 16.6 | 233 | 22 | AAV25524 |
| 18 | 110.5 | 16.6 | 384 | 22 | AAV05479 |
| 19 | 109.5 | 16.5 | 1193 | 17 | AAV05835 |
| 20 | 109.5 | 16.5 | 1193 | 21 | AAV59599 |
| 21 | 109 | 16.4 | 124 | 21 | AAV82334 |
| 22 | 107.5 | 16.2 | 2703 | 22 | ABV60266 |
| 23 | 107 | 16.1 | 250 | 22 | AAV79002 |
| 24 | 107 | 16.1 | 301 | 22 | AAV79986 |
| 25 | 107 | 16.1 | 1712 | 13 | AAV22461 |
| 26 | 106 | 16.0 | 229 | 22 | AAV79404 |
| 27 | 105.5 | 15.9 | 1055 | 19 | AAV44298 |
| 28 | 105.5 | 15.9 | 1148 | 20 | AAV87895 |
| 29 | 105.5 | 15.9 | 1212 | 19 | AAV44299 |
| 30 | 105.5 | 15.9 | 1257 | 17 | AAV05834 |
| 31 | 105.5 | 15.9 | 1257 | 11 | AAV59598 |
| 32 | 105 | 15.8 | 727 | 11 | AAV05533 |
| 33 | 105 | 15.8 | 762 | 21 | AAV92455 |
| 34 | 105 | 15.8 | 3680 | 22 | ABV70878 |
| 35 | 104 | 15.7 | 154 | 22 | AAV78420 |
| 36 | 104 | 15.7 | 167 | 22 | AAV39466 |
| 37 | 104 | 15.7 | 222 | 22 | AAV41252 |
| 38 | 103.5 | 15.6 | 787 | 18 | AAV13573 |
| 39 | 103 | 15.5 | 718 | 11 | AAV05936 |
| 40 | 103 | 15.5 | 762 | 21 | AAV92448 |
| 41 | 103 | 15.5 | 762 | 21 | AAV92449 |
| 42 | 103 | 15.5 | 762 | 21 | AAV92450 |
| 43 | 103 | 15.5 | 762 | 21 | AAV92451 |
| 44 | 103 | 15.5 | 762 | 21 | AAV92452 |
| 45 | 103 | 15.5 | 762 | 21 | AAV92453 |

ALIGNMENTS

| | | |
|----------|--|----------------------------|
| RESULT 1 | AAV09278 | standard; Protein; 112 AA. |
| ID | AAV09278: | |
| AC | AAV09278: | |
| XX | 06-JUL-1999 | (first entry) |
| XX | | |
| DE | YL-1 thermal hysteresis protein. | |
| XX | | |
| KW | Antifreeze; thermal hysteresis protein; THP; mealworm beetle; Tenebrio molitor; absorption-inhibition; ice crystal; frozen food. | |
| XX | | |
| OS | Tenebrio molitor. | |
| XX | | |
| PN | WO9900493-A1. | |
| XX | | |
| PD | 07-JAN-1999. | |
| XX | | |
| PF | 25-JUN-1998; | 98WO-CA00618. |
| XX | | |
| PR | 26-JUN-1997; | 97US-0882907. |
| XX | | |
| PA | (T00H) UNIV QUEBENS KINGSTON. | |
| XX | | |
| PI | Davies PL, Graham JA, Liou Y, Walker VK; | |
| XX | | |
| DR | WPI, 1999-095739/08. | |
| XX | | |
| DR | N-PSDB; AAX33465. | |
| XX | | |
| PT | New nucleic acid encoding antifreeze proteins of Tenebrio molitor - | |
| XX | used to improve low temperature tolerance of fish, plants etc., and | |
| PT | quality of foods or biological materials during frozen storage | |
| XX | | |

Applicant's
own work.

DR N-PSDB: AAX33466.

XX New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PR quality of foods or biological materials during frozen storage

PS Disclosure: Page 62; 88pp; English.

XX The present invention describes a nucleic acid (I) encoding an antifreeze
CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c)
CC (i) binds specifically to antibodies raised against the Tenebrio molitor
CC YL1-YL4 antifreeze proteins, (ii) is at least 60% identical with YL1-4
CC or (iii) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXAXCXAXT (1). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically
CC to reduce formation of ice crystals following exposure to repeated
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more
CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (II) is usual immunoassays. (II) from T. molitor have specific
CC activity 100 times greater than that of fish antifreeze proteins. The
CC present sequence represents a thermal hysteresis protein (THP) from the
CC present invention.

XX Sequence 112 AA:

Query Match 90.2%; Score 599; DB 20; Length 112;
Best Local Similarity 90.2%; Pred. No. 2.1e-40;
Matches 101; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 1 MAFKTCGSKMLVIAIVMCLCTECYCHCTGADCTSCDTACTGCGNCPNATCTDSKN 60
DB 1 MAFKTCGSKMLVIAIVMCLCTECYCHCTGADCTSCDTACTGCGNCPNATCTDSKN 60
OY 61 CVKATCTGTCSTKCNATFCTCTNSKDCFEAKTCTDSTNCKYKATACSTNGCGPGH 112
DB 61 CVKATCTGTCSTKCNATFCTCTNSKDCFEAKTCTDSTNCKYKATACSTNGCGPGH 112

RESULT 4
AAY09282
ID AAY09282 standard; Protein: 124 AA.

XX AAY09282;

XX 06-JUL-1999 (first entry)

XX 5-15 thermal hysteresis protein.

XX Antifreeze: thermal hysteresis protein; THP: mealworm beetle;
KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.

XX Tenebrio molitor.

XX WO9900493-A1.

XX 07-JAN-1999.

XX 25-JUN-1998; 98WO-CA00618.

XX 26-JUN-1997; 97US-0882907.

XX (TOOH) UNIV QUEENS KINGSTON.

XX Davies PL, Graham LA, Liou Y, Walker VK;

DR WPI: 1999-095739/08.

DR N-PSDB: AAX33469.

XX New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PR quality of foods or biological materials during frozen storage

PS Disclosure: Page 67-68; 88pp; English.

XX The present invention describes a nucleic acid (I) encoding an antifreeze
CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c)
CC (i) binds specifically to antibodies raised against the Tenebrio molitor
CC YL1-YL4 antifreeze proteins, (ii) is at least 60% identical with YL1-4
CC or (iii) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXAXCXAXT (1). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically
CC to reduce formation of ice crystals following exposure to repeated
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more
CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (II) is usual immunoassays. (II) from T. molitor have specific
CC activity 100 times greater than that of fish antifreeze proteins. The
CC present sequence represents a thermal hysteresis protein (THP) from the
CC present invention.

XX Sequence 124 AA:

Query Match 90.2%; Score 599; DB 20; Length 124;
Best Local Similarity 83.1%; Pred. No. 2.3e-40;
Matches 103; Conservative 3; Mismatches 6; Indels 12; Gaps 1;

OY 1 MAFKTCGSKMLVIAIVMCLCTECYCHCTGADCTSCDTACTGCGNCPNATCTDSKN 53
DB 1 MAFKTCGSKMLVIAIVMCLCTECYCHCTGADCTSCDTACTGCGNCPNATCTDSKN 60
OY 54 -----TCTDSKNCKVKAATCTGTCSTKCNATFCTCTNSKDCFEAKTCTDSTNCKYKATACSTNG 108
DB 61 CVKATCTGTCSTKCNATFCTCTNSKDCFEAKTCTDSTNCKYKATACSTNG 120

OY 109 CPGH 112

DB 121 CPGH 124

RESULT 5
AAY09280
ID AAY09280 standard; Protein: 148 AA.

XX AAY09280;

XX 06-JUL-1999 (first entry)

XX YL-4 thermal hysteresis protein.

XX Antifreeze: thermal hysteresis protein; THP: mealworm beetle;
KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.

XX Tenebrio molitor.

XX WO9900493-A1.

XX 07-JAN-1999.

XX 25-JUN-1998; 98WO-CA00618.

XX 26-JUN-1997; 97US-0882907.

Best Local Similarity 36.3%; Pred. No. 0.00019;
Matches 37; Conservative 1; Mismatches 44; Indels 20; Gaps 6;

OY 23 CTECHCTGADCTCTDADCTGCG-----CNCNPHCTDCKNVKAAAT-----CTGSTR 72
Db 1118 CTTCCCTGTGGCGCTACG-ACGGCCCTGCTGGTGGCCAGTACAGATGGACACAGCTG 1176
OY 73 CNTARTCTNSKDCFEAKTCTDSTNCYKATACCTNSTG--CPG 111
Db 1177 C-----TCCGGAGCCGTGACTGATC--AACCAAGTGTACCTG 1212

RESULT 9
AAID 74791
ID AAY74791 standard; Protein: 233 AA.

AC AAY74791;
XX
DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 263 protein sequence SEQ ID NO:1056.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.

OS Neisseria meningitidis.

PN WO957280-A2.

XX 11-NOV-1999.

PD 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.

PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;

DR WPI; 2000-062150/05.
DR N-PSDB; AAZ53553.

PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -

PS Claim 2; Page 606; 1453pp; English.

XX AA53015 to AA54536, AA54537 to AA54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA54537 to AA54576 and AA54616 to AA55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX Sequence 233 AA;

Query Match 20.9%; Score 139; DB 21; Length 233;

Best Local Similarity 38.7%; Pred. No. 0.00074;
Matches 36; Conservative 3; Mismatches 42; Indels 12; Gaps 7;

OY 21 CLETCCHCTG-GADCTCTDADCTGCGNCPNPHCTDCKNVKAAATC-TGST--KCMYA 76
Db 101 CATTTCGATGGGGCTCAACCAACGACAGC--AGGCAGI--CGAAGTGTATGGGCGTA 156

OY 77 RCTCTNSKDCFEAKTCTDSTNCYKATACCTNSTGC 109

Db 157 SYCTGG--C--AACCCCTGTGCAACTACGTCAAC 185

RESULT 10
AAID 9276
ID AAY09276 standard; peptide; 24 AA.

AC AAY09276;

DT 06-JUL-1999 (first entry)

DE Antifreeze protein amino acid sequence SEQ ID NO:4.

KW Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
KM Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.

OS Synthetic.
OS Tenebrio molitor.

FN Key Location/Qualifiers

FT Misc-difference 1..24
FT FT
XX /note= "X is unspecified"

PN WO9900493-A1.

XX 07-JAN-1999.

PD 25-JUN-1998; 98WO-CA00618.

PF 26-JUN-1997; 97US-0882907.

PR (TOOH) UNIV QUEBENS KINGSTON.

PA Davies PL, Graham LA, Liou Y, Walker VK;

PI WPI; 1999-095739/08.

PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PT quality of foods or biological materials during frozen storage

PS Claim 14; Page 56; 88pp; English.

XX The present invention describes a nucleic acid (I) encoding an antifreeze
CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c)
CC (1) binds specifically to antibodies raised against the Tenebrio molitor
CC YLI-YLA antifreeze proteins, (11) is at least 60% identical with YLI-4
CC or (111) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXSKXCKXKXMT (I). (11), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically
CC to reduce formation of ice crystals following exposure to repeated
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more
CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (11) is usual immunoassays. (11) from T. molitor have specific

[illegible]

Job time : 45.6471 secs

```
OY      5  TCGFSKKMLVIAIVIMC-----LCTEYCHCTGADCTISCTDA-----CTGC 46
      11  :      |      |      |      |      |      |      |      |
Db      636  TCAAAAGCAACATATCCAAATGATAACATCTGAGAGATCAACAGAAATTAATCCATGC 695
OY      47  GNCPPAHTCTDSKNCVKAATCTGSKCNTARTC-----TNSKDCFBAKTCYDSTNCY--- 98
      11  :      |      |      |      |      |      |      |      |
Db      696  -----AAACT---TC--AATATGATTAATACTAATCTTACTCATGTGTTTGGTCAAGA 745
OY      99  -KATACTNSTG---CPG 111
      11  :      |      |      |      |      |      |      |
Db      746  GTATCCTTGTGAACGG 762
```

RESULT 15

AAB36084
ID AAB36084 standard; Protein: 225 AA.

AC AAB36084;

DT 16-FEB-2001 (first entry)

DE Neospora caninum NCP20 antigen.

XX Neospora caninum; NCP20; antiparasitic; vaccine; antigen;

KW neosporosis.

OS Neospora caninum.

PN WO20063244-A1.

PD 26-OCT-2000.

PF 20-APR-2000; 2000WO-AU00354.

PR 21-APR-1999; 99AU-000928.

PA (INSE-) INSEARCH LTD.

PI Ellis JT, Atkinson R, Ryce CL;

DR WPI; 2000-679588/66.

DR N-PSDB; AAC67872.

PT Polynucleotide encoding parasite antigens and the antigenic polypeptide
PT useful for immunizing mammals against neosporosis -

PS Claim 4; Page 35-36; 39pp; English.

XX
XX
CC The present sequence is Neospora caninum antigen NCP20. A
CC composition comprising NCP20 is useful for obtaining a protective effect
CC against neosporosis in a mammal. The NCP20 polynucleotide may be
CC administered in the form of a DNA vaccine for vaccinating animals against
CC neosporosis. Antibodies against the NCP20 antigen may be identified
CC using the NCP20 polypeptide. NCP20 and antibodies specific to NCP20 are
CC useful for diagnosis of neosporosis.

SQ Sequence 225 AA;

Query Match 16.7%; Score 111; DB 21; Length 225;

Best Local Similarity 32.5%; Pred. No. 0.11;

Matches 27; Conservative 3; Mismatches 43; Indels 10; Gaps 3;

OY 30 CTGGADCTCTDAGTCGCGNCPAHTCTDSKNCVKAATCTGSKCN-TARTCTNSKDCFEA 88

DB 18 CTTGTGCGCAAAACGACGCC-----CCTCCGCGTACCCGTTTGGCTAGTCT---CTAT 68

OY 89 KTCTDSTNCKYKATCTNSTGCPG 111

DB 69 TCCAGTAGCGCGTTCTGAATCCG 91

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:37:29 ; Search time 17.2941 Seconds
(without alignments)
158.185 Million cell updates/sec

Title: US-10-032-658-11

Perfect score: 664

Sequence: 1 MARCTGCFSKKMLVIAIVM.....DSTNCKYKATACNTSGCPGH 112

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 377.5 | 56.9 | 108 | 1 | US-08-485-359-2 |
| 2 | 377.5 | 56.9 | 108 | 1 | US-08-569-594-2 |
| 3 | 377.5 | 56.9 | 108 | 5 | PCT-US96-08815-2 |
| 4 | 361 | 54.4 | 109 | 1 | US-08-485-359-4 |
| 5 | 361 | 54.4 | 109 | 1 | US-08-569-594-4 |
| 6 | 361 | 54.4 | 109 | 5 | PCT-US96-08815-4 |
| 7 | 160.5 | 24.2 | 1345 | 2 | US-08-977-767-3 |
| 8 | 159.5 | 24.0 | 120 | 3 | US-08-508-761B-22 |
| 9 | 157.5 | 23.7 | 1417 | 4 | US-08-900-230-3 |
| 10 | 139 | 20.9 | 1400 | 4 | US-08-630-915A-37 |
| 11 | 130.5 | 19.7 | 801 | 1 | US-07-906-349A-6 |
| 12 | 117 | 17.6 | 143 | 4 | US-08-990-823-112 |
| 13 | 115.5 | 17.4 | 102 | 3 | US-08-974-022-53 |
| 14 | 115.5 | 17.4 | 102 | 4 | US-08-795-445A-53 |
| 15 | 115.5 | 17.4 | 102 | 4 | US-08-795-447A-53 |
| 16 | 115.5 | 17.4 | 102 | 4 | US-08-974-186-53 |
| 17 | 115.5 | 17.4 | 102 | 4 | US-08-795-446B-53 |
| 18 | 114 | 17.2 | 2556 | 1 | US-08-083-590A-20 |
| 19 | 114 | 17.2 | 2556 | 3 | US-08-532-384-20 |
| 20 | 113 | 17.0 | 341 | 2 | US-08-209-521-11 |
| 21 | 113 | 17.0 | 2523 | 1 | US-08-185-432-18 |
| 22 | 112 | 16.9 | 2556 | 1 | US-08-185-432-17 |
| 23 | 111 | 16.7 | 1248 | 4 | US-08-882-046-6 |
| 24 | 109.5 | 16.5 | 1193 | 3 | US-08-400-158-10 |
| 25 | 109.5 | 16.5 | 1193 | 2 | US-08-611-728A-10 |
| 26 | 109.5 | 16.5 | 2703 | 4 | US-08-185-432-19 |
| 27 | 107.5 | 16.2 | 45 | 1 | US-08-900-230-14 |

| | | | | | | |
|----|-------|------|------|---|-------------------|--------------------|
| 28 | 106.5 | 16.0 | 1219 | 4 | US-08-882-046-5 | Sequence 5, Appli |
| 29 | 105.5 | 15.9 | 47 | 3 | US-08-482-085B-91 | Sequence 91, Appl |
| 30 | 105.5 | 15.9 | 1055 | 4 | US-09-214-278-2 | Sequence 2, Appli |
| 31 | 105.5 | 15.9 | 1065 | 2 | US-08-400-159-8 | Sequence 8, Appli |
| 32 | 105.5 | 15.9 | 1148 | 4 | US-08-882-046-4 | Sequence 4, Appli |
| 33 | 105.5 | 15.9 | 1212 | 4 | US-09-214-278-3 | Sequence 3, Appli |
| 34 | 105.5 | 15.9 | 1257 | 4 | US-08-611-729A-8 | Sequence 8, Appli |
| 35 | 105 | 15.8 | 726 | 6 | 5208144-37 | Patent No. 5208144 |
| 36 | 104.5 | 15.7 | 1338 | 4 | US-09-214-278-5 | Sequence 5, Appli |
| 37 | 103 | 15.5 | 718 | 1 | US-08-444-792-4 | Sequence 4, Appli |
| 38 | 103 | 15.5 | 718 | 1 | US-08-445-042-4 | Sequence 4, Appli |
| 39 | 103 | 15.5 | 788 | 2 | US-07-728-215-32 | Sequence 32, Appl |
| 40 | 103 | 15.5 | 788 | 4 | US-08-938-085A-32 | Sequence 32, Appl |
| 41 | 102.5 | 15.4 | 577 | 2 | US-07-728-215-29 | Sequence 29, Appl |
| 42 | 102.5 | 15.4 | 577 | 4 | US-08-938-085A-29 | Sequence 29, Appl |
| 43 | 102 | 15.4 | 799 | 2 | US-08-525-940-23 | Sequence 23, Appl |
| 44 | 102 | 15.4 | 799 | 2 | US-08-976-838-23 | Sequence 23, Appl |
| 45 | 102 | 15.4 | 881 | 2 | US-08-525-940-21 | Sequence 21, Appl |

ALIGNMENTS

```

RESULT 1
US-08-485-359-2
; Sequence 2, Application US/08485359
; Patent No. 5627051
; GENERAL INFORMATION:
; APPLICANT: Duman, John G.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
; TITLE OF INVENTION: DENDROIDES CANADENSIS ANTIREPZE PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 South Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,359
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lammert, Steven R.
; REGISTRATION NUMBER: 27653
; REFERENCE/DOCKET NUMBER: 835910-25377
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7258
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ. ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Dendroides canadensis
; US-08-485-359-2

Query Match 56.9%; Score 377.5; DB 1; Length 108;
Best Local Similarity 64.6%; Pred. No. 7.5e-27;
Matches 64; Conservative 13; Mismatches 21; Indels 1; Gaps 1;
13 LVIAIVMCLTECYCHCTGAGADCTCTDACTGCGNPNMHT-CTDSKNCVKATCTGSR 71
||||:|:| ||| ||||| | | ||||| | | | ||| |

```


Sequence 3, Application US/08977767
Patent No. 5972684
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Yue, Henry
APPLICANT: Greenwald, Sara
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,767
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0423 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1345 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1532042
US-08-977-767-3

Query Match 24.2%; Score 160.5; DB 2: Length 1345;
Best Local Similarity 38.3%; Pred. No. 9e-07;
Matches 41; Conservative 2; Mismatches 45; Indels 19; Gaps 6;

QY 16 AVTVMCLCTECYCHCTGADCTSC--TDACTGCGNCPNAHTCTDSKNCVKAA---TCTGS 70
DB 1012 AATATGACCTCCCTCCAG--CTGCATTCCTGTGAGACT---TGTGAGCCAGATCTCTCGG 1066

QY 71 TKCNTARTCTNSKDCFEATCTDSTNCTYAT-----ACTNSTGCPG 111
DB 1067 GTCCAGCTCT---CCAGAGCTCAGCCGTAAAGCGCCGCTCGAG 1110

RESULT 8
US-08-508-761B-22
Sequence 22, Application US/08508761B
GENERAL INFORMATION:
APPLICANT: Joliff, Gwennael
APPLICANT: Guyonvarch, Arnel
APPLICANT: Purification, Relano
APPLICANT: Duchilton, Francis
APPLICANT: Renaud, Michel
TITLE OF INVENTION: System for Protein Expression and
TITLE OF INVENTION: Secretion Especially in Corynebacteria
NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
STREET: 400 Seventh St. N.W.
CITY: Washington D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,761B
FILING DATE: 31-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09652
FILING DATE: 29-JUL-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09870
FILING DATE: 02-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,409
REFERENCE/DOCKET NUMBER: P58525NA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 393-5350
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 120 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: synthetic
US-08-508-761B-22

Query Match 24.0%; Score 159.5; DB 3: Length 120;
Best Local Similarity 39.1%; Pred. No. 1.3e-07;
Matches 34; Conservative 1; Mismatches 49; Indels 3; Gaps 3;

QY 23 CTCTCYCHCTGADCTSCDTACTGCGNCPNAHTCTDSKNCVKAAATCTGCTCNTARTCTNS 82
DB 36 CTGCGC-CTGGGCGCTGC-GCCTGTGCTGAGACCTGGGCGCTGGGCGCTGGGCGCTGT 93

QY 83 KDCFEAKTCTDSTNCTYATCTACTNSTGC 109
DB 94 GCTGGGCGCT-GAGCCTGTGCTGCTGTC 119

RESULT 9
US-08-900-230-3
Sequence 3, Application US/08900230
Patent No. 6329197
GENERAL INFORMATION:
APPLICANT: Baird, Jonathan A.
TITLE OF INVENTION: DNA ENCODING GALANN GALR3 RECEPTORS AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of The Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 11036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/900,230
FILING DATE: 23-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-900-230-3

Query Match 23.7%; Score 157.5; DB 4; Length 1417;
Best Local Similarity 37.1%; Pred. No. 1.7e-06;
Matches 36; Conservative 0; Mismatches 50; Indels 11; Gaps 3;

QY 21 CLCTEYCHCT-----GGADCTCTDCTGCGNCPNHHCTDSKNCVKAATCTGTCN 74
DB 904 CTCACGGGCTCTGCTGGGTCGCC--ACGACGGCTCATCTCTGC---TTCGTGACG 958
QY 75 TARTCTNSKDCFEAKTCTDSTNCKRATCTNCTGCPG 111
DB 959 GCCGCTTCGCTTCAGCCGCCGACCTACGCTGTCGC 995

RESULT 10
US-08-630-915A-37
Sequence 37, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOPMAN, NO. 6309820H
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:
LENGTH: 1400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-37

Query Match 20.9%; Score 139; DB 4; Length 1400;
Best Local Similarity 29.7%; Pred. No. 7.3e-05;
Matches 38; Conservative 3; Mismatches 53; Indels 34; Gaps 5;

QY 16 AVIWMCLC-----TECYCHC-----TGADCTCTDCTGCGNCPNHHCTDSK 59
DB 347 AAATCCAGAAATAGAGTTCCCGTCAGAGGAACCACTGATTCACATCTGCC 406
QY 60 NCVK-----AATCTG---STKCNFAR-----TCTNSKDCFEAKTCTDSTNCTYAT 101
DB 407 CCTGCCCCCAACTGGCTTGGTGAGACCCGCCCTTGGCAGTAACTCTTCAGAG 466
QY 102 ACTNSTGC 109
DB 467 CCCTCCAC 474

RESULT 11
US-07-906-349A-6
Sequence 6, Application US/07906349A
Patent No. 5434064
GENERAL INFORMATION:
APPLICANT: Schlessinger, Joseph
APPLICANT: Skolnik, Edward Y.
APPLICANT: Margolis, Benjamin L.
TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASES
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Broadway and Neimark
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/906,349A
FILING DATE: 30-JUN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/643,237
FILING DATE: 18-JAN-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-906-349A-6
Query Match 19.7%; Score 130.5; DB 1; Length 801;
Best Local Similarity 35.1%; Pred. No. 0.00025;
Matches 34; Conservative 3; Mismatches 39; Indels 21; Gaps 6;
QY 23 CTETCYCHCTG-----GADCTCTDCTGCGNCPNHHCTDSKNCVKAATCTGTCNCTG 77

Db 556 CTTCACAAATAATGTTTAATCATCATGCTCC-----TC-----CTCTTCCCTCCCTTTG 606
QY 78 TCTNSKDCFEAKTC--IDSTNCYKATACTNSTG-CPG 111
Db 607 TCTTTTTCATCCTTTTTC-----TCTTCTGTCAG 639

RESULT 12

US-08-990-823-112
; Sequence 112, Application US/08990823D
; Patent No. 6228371
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 49086
; CURRENT APPLICATION NUMBER: US/08/990,823D
; EARLIER FILING DATE: 1997-12-15
; EARLIER APPLICATION NUMBER: US 96/10375
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: 60/000,254
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 112
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-08-990-823-112

Query Match

Best Local Similarity 17.6%; Score 117; DB 4; Length 143;
Matches 36; Conservative 2; Mismatches 47; Indels 40; Gaps 8;

QY 21 CICTEC-----YC--HCTGADC-----TCTACTGCG-----NCPNAHTCTD 57
Db 17 CACACACAGTAGTCAGCAGAGACCCCGCCAGCTTGGCGCGGAGATCTTCCCTG 76
QY 58 SKNCVKATCTGCTKCTNT-----ARTCTNSKDCFEAKTCTDSTNCYKATACTN-----S 106
Db 77 TGCG-----TCACCGACTTTCGCCGCAACTTCAC--GATCTTGCAATGAGCGGCAAGAA 130
QY 107 TGCPG 111
Db 131 CGCGC 135

RESULT 13

US-08-974-022-53
; Sequence 53, Application US/08974022
; Patent No. 6015938
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,022

; FILING DATE: 12-DEC-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-974-022-53

Query Match

Best Local Similarity 17.4%; Score 115.5; DB 3; Length 102;
Matches 35; Conservative 2; Mismatches 38; Indels 25; Gaps 7;

QY 23 CTECHCHCTGGADCTISCTDACTG-----CGNCPNAHTCTDSKNCVKATC--TGST 71
Db 15 CTCTCATCAG--CTGCTGTGTGATTAATGTCGCCGGGTACA--CCTACTCTTTGAA 69
QY 72 KCTARTCTNSKDCFEAKTCTDSTNCYKATACTNSTGCGP 111
Db 70 GAGTAGTCGACGAC--ACACTATT-----TACAG--GC GG 100

RESULT 14

US-08-795-445A-53
; Sequence 53, Application US/08795445A
; Patent No. 6284485
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,445A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-445A-53

Query Match

Best Local Similarity 17.4%; Score 115.5; DB 4; Length 102;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:36:38 ; Search time 22.2353 seconds
(without alignments)
484.006 Million cell updates/sec

Title: US-10-032-658-11

Perfect score: 664
Sequence: 1 MAFKTCGFSSKMWLVIAIVM.....DSTNCYKATACNSTGCPGH 112

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 138.5 | 20.9 | 313 | 2 | extracellular matr |
| 2 | 128.5 | 19.4 | 1046 | 2 | prestalk protein p |
| 3 | 119.5 | 18.0 | 861 | 2 | Notch homolog Motc |
| 4 | 118 | 17.8 | 4660 | 2 | T42737 |
| 5 | 117.5 | 17.7 | 2531 | 2 | S18188 |
| 6 | 117.5 | 17.7 | 2531 | 2 | A46019 |
| 7 | 114 | 17.2 | 2555 | 2 | A40043 |
| 8 | 113.5 | 17.1 | 548 | 2 | T16642 |
| 9 | 113.5 | 17.1 | 3002 | 2 | A47221 |
| 10 | 113 | 17.0 | 2524 | 2 | A35844 |
| 11 | 112.5 | 16.9 | 169 | 1 | S18946 |
| 12 | 112 | 16.9 | 1111 | 2 | T26972 |
| 13 | 111 | 16.7 | 677 | 2 | C42125 |
| 14 | 111 | 16.7 | 1101 | 2 | T16840 |
| 15 | 110.5 | 16.6 | 2871 | 2 | A55567 |
| 16 | 109 | 16.4 | 230 | 2 | A38346 |
| 17 | 109 | 16.4 | 798 | 2 | A40526 |
| 18 | 108.5 | 16.3 | 640 | 1 | A55624 |
| 19 | 108 | 16.3 | 440 | 1 | A30452 |
| 20 | 107.5 | 16.2 | 188 | 2 | JC6547 |
| 21 | 107.5 | 16.2 | 1700 | 2 | S08167 |
| 22 | 107.5 | 16.2 | 2703 | 1 | A24420 |
| 23 | 107 | 16.1 | 601 | 2 | A27020 |
| 24 | 107 | 16.1 | 1712 | 2 | A38261 |
| 25 | 107 | 16.1 | 2352 | 2 | T30201 |
| 26 | 106.5 | 16.0 | 1220 | 2 | A56136 |
| 27 | 106.5 | 16.0 | 3507 | 2 | T34513 |
| 28 | 106 | 16.0 | 223 | 2 | B38346 |
| 29 | 106 | 16.0 | 738 | 2 | S40992 |

| | | | | | | |
|----|-------|------|------|---|--------|--------------------|
| 30 | 106 | 16.0 | 739 | 2 | B88553 | protein K04H4.2b [|
| 31 | 105.5 | 15.9 | 574 | 2 | B88465 | protein B0244.8 [1 |
| 32 | 104 | 15.7 | 674 | 2 | I55476 | growth potentiati |
| 33 | 104 | 15.7 | 2825 | 2 | T14271 | Doc4 protein, stre |
| 34 | 103.5 | 15.6 | 58 | 2 | S43367 | metallothionein - |
| 35 | 103.5 | 15.6 | 65 | 2 | A38739 | integrin beta-3 ch |
| 36 | 103.5 | 15.6 | 680 | 2 | PN0510 | furin (EC 3.4.21.7 |
| 37 | 103.5 | 15.6 | 1299 | 2 | T43251 | hypothetical prote |
| 38 | 103.5 | 15.6 | 1895 | 2 | T15881 | notch homolog - se |
| 39 | 103.5 | 15.6 | 2531 | 2 | T31070 | platelet glycoprot |
| 40 | 103 | 15.5 | 753 | 2 | B36268 | platelet glycoprot |
| 41 | 103 | 15.5 | 778 | 2 | A60798 | platelet glycoprot |
| 42 | 103 | 15.5 | 788 | 2 | A26547 | platelet glycoprot |
| 43 | 103 | 15.5 | 788 | 2 | I77349 | platelet glycoprot |
| 44 | 103 | 15.5 | 1557 | 2 | T28811 | hypothetical prote |
| 45 | 103 | 15.5 | 2946 | 2 | T15840 | hypothetical prote |

ALIGNMENTS

RESULT 1
S44208
extracellular matrix protein B - Dictyostelium minutum (fragment)
C:Species: Dictyostelium minutum
C:Date: 07-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 05-Nov-1999
C:Accession: S44208
Ryan Es, S.; Nieuwenhuisen, B.W.; Lenouvel, F.; van Deursen, E.M.; Schaap, P.
Submitted to the EMBL Data Library, April 1994
A:Description: Universal signals control slime mold stalk information.
A:Reference number: S44208
A:Accession: S44208
A:Molecule type: DNA
A:Residues: 1-313 <V>
A:Cross-references: EMBL:X78948; NID:9474918; PIDN:CAA55545.1; PID:9474919
A:Genetics:
A:Introns: 24/1
C:Keywords: extracellular matrix

Query Match 20.9%; Score 138.5; DB 2; Length 313;
Best Local Similarity 29.6%; Pred. No. 0.0015;
Matches 32; Conservative 9; Mismatches 46; Indels 21; Gaps 3;
QY 23 CTCTCTCTGGA-----DCTCT-DACTGCGCPNAHTCTPSKNCVKATCTGSKTC 73
DB 91 CLTDCSDCKGCGCTHPMCDHNCTLDSCDCKGCGCTHPMCDCKNACTVDSGNSSTGC 150
QY 74 N-----TARTCTNSKDCFEAKTCPTDSTNCYKATACNSTGCG 109
DB 151 SHTPTSCDYNSTVDSCDCKGCGVHTPTACDKNACTVDSGNSKTCG 198
RESULT 2
A26838
prestalk protein precursor - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 22-Nov-1996
C:Accession: A26838
R:Ceccarelli, A.; McRobbie, S.J.; Jernyn, K.A.; Duffy, K.; Early, A.; Williams, J.G.
Nucleic Acids Res. 15; 7463-7476, 1987
A:Title: Structural and functional characterization of a Dictyostelium gene encoding
A:Reference number: A26838; MUID:88015608
A:Accession: A26838
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1046 <CEC>
A:Experimental source: Strain Ax-2
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1046/Product: prestalk protein #status predicted <MAT>
Query Match 19.4%; Score 128.5; DB 2; Length 1046;
Best Local Similarity 35.9%; Pred. No. 0.015;
Matches 33; Conservative 9; Mismatches 37; Indels 13; Gaps 5;

| | | | | | | | | | | |
|----|-----|-----------------|----|----------|---------|--------------|---------------|----------|-------------|-----|
| OY | 23 | CTTCYCHCTGAGD | -- | SCD | - | ACGCGCNC | PAHCTCTSKNCYA | - | ATCGTCTCNAR | 77 |
| | | | | | | | | | | |
| Db | 800 | CTEKCCTQNGSVYHT | PI | RCDLNSCI | ----- | ADCSNSTGCVHT | PI | ICDDNNKC | -TAD | 851 |
| | | | | | | | | | | |
| OY | 78 | TCTNSKOCFEAKTCT | DS | TCYKAT | ACTNSG | CTGC | | | | 109 |
| | | | | | | | | | | |
| Db | 852 | SCSNSTGCGHTPI | SC | DDNNPCTV | SCSNSTG | CTGC | | | | 883 |
| | | | | | | | | | | |

RESULT 3

Notch homolog Notch protein - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
 C:Accession: A48825
 C:Reanne, A.G.; Conlon, R.A.; Ziringibl, R.; Yamaguchi, T.P.; Rossant, J.
 Dev. Biol. 154, 377-387, 1992
 A:Title: Expression analysis of a Notch homologue in the mouse embryo.
 A:Reference number: A48825; MUID:93050801
 A:Accession: A48825
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-861 <RG>
 A:Experimental source: embryo
 A:Note: sequence extracted from NCBI backbone (NCBIP:119144)
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology;
 F:26-57/Domain: EGF homology <EGF>
 F:64-95/Domain: EGF homology <EGX1>
 F:199-229/Domain: EGF homology <EGF2>
 F:441-472/Domain: EGF homology <EGX2>

| Query Match | 18.0% | Score 119.5; | DB 2; | Length 861; |
|-------------|-------|--------------|-------|-------------|
| | | | | |

Matches 36; Conservative 7; Mismatches 50; Indels 17; Gaps 6;

QY 15 IAVLYMC-----LCTE----CYCHCTGGADCTSCIDACTGC--GNCENAHCTCID--SKNC 61
| | ::| | : | | | | : | | | | : |
Db 318 IDVTLCCGSGLCVDECDKHVHCQAGVTSYCEDEVEDECSNPQNGATCTDYLGRSNC 377

QY 62 VKATCTGSTKNTARTCTNSKDCFEAKTCTDSTNCYKATACTNSTGCPG 111

RESULT 4

gp330 protein precursor - rat
N:Alternate names: megalin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jan-2000 #sequence
C:Date: 11-Jan-2000 #text-change 04-Mar-2000
C:Accession: I42737
R:Salto, A.; Pletromonaco, S.; Loo, A.K.C.; Fargunhar, M.G.
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994
A:Title: Complete cloning and sequencing of rat gp330/megalyn, a distinctive member of the gp330 protein family
A:Reference number: A58173; MUID:95024033
A:Accession: T42737
A:Status: preliminary; translated from GR/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-4660 <SAIT>
A:Cross-references: EMBL:L34049; NID:9561852; PID:9561853; PIDN:AAA51369.1
A:Experimental source: Strain Sprague-Dawley; kidney
A:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
A:1-25/Domain: signal sequence #status predicted <SIG>
A:26-4660/Product: gp330 protein #status predicted <MAT>

| | | | | |
|-----------------------|--------|-----------------|-------|----------------|
| Query Match | 17.8%; | Score 118; | DB 2; | Length 4660; |
| Best Local Similarity | 23.7%; | Pred. No. 0.21; | | |
| Matches | 36; | Conservative | 14; | Mismatches 46; |
| | | | | Indels 56; |
| | | | | Gaps 8; |

```

0y 12 WLVAIVIVMCL-----CTECYC---HCTGADCTCTCD----- 41
      |::| | | | | |
Db 10 WMLLIAIACLEPVSSQECGSGNRCNDNGYICIPASMRKCDGTRCDLDDTDEIGCPRSCS 65

```

```

OY 42 ---ACTGGGCN-PAHAACTGDKNNVKA-----TCGSGKCTAATCTNSK-----DCFE 87
Db 70 GLFLCPACGTCTIPSSWCYDEKDCSDGADBEQGNAGTTCGACMTGCSNGCCISPEYICDH 129
OY 88 AKTC---TDSNCKYKAT-----ACTNST 107
Db 130 VSDCPDGSDBERNCHPPTCDQLTCANGCACTYNTS 161

```

RESULT 5

notch protein homolog - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C:Accession: S18188
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A:Title: A homolog of Drosophila Notch expressed during mammalian development.
A:Reference number: S18188; MUID:92111383
A:Accession: S18188
A:Molecule type: mRNA
A:Residues: 1-2531 <ME1>
A:Cross-references: EMBL:X57405; NID:957634; PID:957635
C:Superfamily: unassigned anktyrin repeat proteins; anktyrin repeat homology; EGF homol
F:1025-1056/Domain: EGF homology <EGF1>
F:1025-1056/Domain: EGF homology <EGF>
F:1233-1264/Domain: EGF homology <EGF2>
F:1917-1949/Domain: anktyrin repeat homology <AN1>
F:1950-1982/Domain: anktyrin repeat homology <AN2>
F:1984-2016/Domain: anktyrin repeat homology <AN3>
F:2017-2049/Domain: anktyrin repeat homology <AN4>
F:2050-2082/Domain: anktyrin repeat homology <AN5>

| | | | | |
|-----------------------|--------|-----------------|-------|--------------|
| Query Match | 17.7%; | Score 117.5; | DB 2; | Length 2531; |
| Best Local Similarity | 32.1%; | Pred. No. 0.16; | | |

| | | | | | | | | | |
|---------|-----|--------------|----|------------|-----|--------|-----|------|----|
| Matches | 35; | Conservative | 7; | Mismatches | 46; | Indels | 21; | Gaps | 6; |
|---------|-----|--------------|----|------------|-----|--------|-----|------|----|

```

Oy  15 IAVIMC-----LCIE----CYCHGCGGADCTCTDAGTC--GNCPPAHTCTD--SKNC 61
      |  ::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db  110 IDVTLCCGGGLCYDEEDKHYCHCQAGYTSYCEDEVEDECSNPCCGATCTDYLGSFSC 1169

```

QY 62 VKATCTGSKCNTARTCTNSKDCF EAKTCTDSTNCYKATACTNSTGCP 110

Db 1170 KCVAGYHGSNCSEINECL-SQPCQNGGTCIDLTNTYKCS-----Cp 1210

RESULT 6

Notch-1 protein - mouse
N:Alternate names: notch protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Sep-1993 *sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A46019; S25144
R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G
Genomics 15, 259-264, 1993
A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog
A:Reference number: A46019; MUID:93194170
A:Accession: A46019
A:Status: not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-2531 <GBL>
A:Cross-references: C1211886; GB:S47228; NID:q288502; PIDN:CAA77941.1; PID:q288503
A:Note: sequence extracted from NCBI backbone (NCBI:P127318)
R:Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.
submitted to the EMBL Data Library, April 1992
A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, sugg
A:Reference number: S25144
A:Accession: S25144
A:Molecule type: mRNA
A:Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>
A:Cross-references: EMBL:Z11886
A:Genetics:

A:Gene: notch-1
 A:Map position: 2
 A>Note: Proximal region of chromosome 2
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:106-138/Domain: EGF homology <EGF1>
 F:144-175/Domain: EGF homology <EGF1>
 F:222-254/Domain: EGF homology <EGF2>
 F:261-292/Domain: EGF homology <EGF2>
 F:339-370/Domain: EGF homology <EGF3>
 F:416-449/Domain: EGF homology <EGF3>
 F:456-487/Domain: EGF homology <EGF4>
 F:494-525/Domain: EGF homology <EGF5>
 F:532-563/Domain: EGF homology <EGF6>
 F:607-638/Domain: EGF homology <EGF7>
 F:682-713/Domain: EGF homology <EGF8>
 F:757-788/Domain: EGF homology <EGF9>
 F:795-826/Domain: EGF homology <EGF10>
 F:873-904/Domain: EGF homology <EGF11>
 F:911-942/Domain: EGF homology <EGF12>
 F:949-980/Domain: EGF homology <EGF13>
 F:987-1018/Domain: EGF homology <EGF14>
 F:1025-1056/Domain: EGF homology <EGF15>
 F:1063-1094/Domain: EGF homology <EGF16>
 F:1149-1180/Domain: EGF homology <EGF17>
 F:1187-1218/Domain: EGF homology <EGF18>
 F:1233-1264/Domain: EGF homology <EGF19>
 F:1352-1383/Domain: EGF homology <EGF19>
 F:1391-1425/Domain: EGF homology <EGF>
 F:1917-1948/Domain: ankyrin repeat homology <AN1>
 F:1949-1981/Domain: ankyrin repeat homology <AN2>
 F:1983-2015/Domain: ankyrin repeat homology <AN3>
 F:2016-2048/Domain: ankyrin repeat homology <AN4>
 F:2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 17.7%; Score 117.5; DB 2; Length 2551;
 Best Local Similarity 32.1%; Pred. No. 0.16;
 Matches 35; Conservative 7; Mismatches 46; Indels 21; Gaps 6;
 QY 15 IAVIWC-----LCTE-----CYCHTGAGADCTCTDAGTCG--GNCNPAHTCTD---SKNC 61
 DB 1110 IDTTLCCQHGGLCVDEGDRKHYCHCAQGYGSCYCEDEVDSCSPNQCNGATCTDYLGFSC 1169
 QY 62 VKAATCGSTKCNARTCTNSKDCFEAKTCTDSTNCYKATCTNSTGCP 110
 DB 1170 KCVAGHGSCSEETNECL-SQPCNGGTCIDLTNSYKCS-----CP 1210

RESULT 7
 A40043
 notch protein homolog TAN-1 precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 13-Aug-1999
 C:Accession: A40043
 R:Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar, Cell 66, 649-661, 1991
 A:Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal
 A:Reference number: A40043; MUID:91347367
 A:Accession: A40043
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 1-2555 <ELL>
 A:Cross-references: GB:M73980
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:261-292/Domain: EGF homology <EGF1>
 F:494-525/Domain: EGF homology <EGF1>
 F:987-1018/Domain: EGF homology <EGF2>
 F:1149-1180/Domain: EGF homology <EGF2>
 F:1187-1218/Domain: EGF homology <EGF3>
 F:1233-1264/Domain: EGF homology <EGF3>
 F:1927-1959/Domain: ankyrin repeat homology <AN1>
 F:1960-1992/Domain: ankyrin repeat homology <AN2>
 F:1994-2026/Domain: ankyrin repeat homology <AN3>
 F:2027-2059/Domain: ankyrin repeat homology <AN4>

F:2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match 17.2%; Score 114; DB 2; Length 2555;
 Best Local Similarity 31.6%; Pred. No. 0.29;
 Matches 31; Conservative 6; Mismatches 45; Indels 16; Gaps 6;
 QY 23 CTE-----CYCHTGAGADCTCTDAGTCG--GNCNPAHTCTDSTNCYKATCT-----GST 71
 DB 922 CTDSINTAFDCILPFGFTGCEEDINDCASDPKRNACND---CVDSTCTCTCPAFSGI 978
 QY 72 KC-NTARTCTNSKDCFEAKTCTDSTNCYKATCTNSTG 108
 DB 979 HCENNTPDCTES-SCFNGTCTCVGINSFTCLCPFGFG 1015

RESULT 8
 T16642
 hypothetical protein R01H2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16642
 R:Babbs, N.
 submitted to the EMBL Data Library, February 1994
 A:Description: The sequence of C. elegans cosmid R01H2.
 A:Reference number: 218553
 A:Accession: T16642
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-548 <BAB>
 A:Cross-references: EMBL:U00035; NID:9458989; PID:9458992; PIDN:AAA50645.1; CESP:R01H
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:R01H2.3
 A:Introns: 18/2; 44/3; 259/3; 495/1

Query Match 17.1%; Score 113.5; DB 2; Length 548;
 Best Local Similarity 25.0%; Pred. No. 0.13;
 Matches 37; Conservative 14; Mismatches 36; Indels 61; Gaps 9;
 QY 19 VMCLCTEYCY-----HCTGAGADCTSC-----TDA-----CTGCNCPN--- 51
 DB 188 LITLTAERTLCGVNCPDGSDEASCRSKSCNDPKSGSDACLPISYKCDGVSDCENESD 247
 QY 52 -----AHTCTDSKNCVKAA-TCTGSTKCN-----TARTCTNSKDCFEAKTCT 92
 DB 248 ESNCKKCKGAAHC--GRNCTKASKVCDGIPDCDDGSDERHCDCCKTSGS-----ERALCE 301

RESULT 9
 A47221
 fibrillin 1 precursor - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 02-Jun-1995 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000
 C:Accession: A47221; I54355; S17064; I59574; S17062; S62111; A34198
 R:Corson, G.M.; Chalberg, S.C.; Dietz, H.C.; Charbonneau, N.L.; Sakai, L.Y.
 Genomics 17, 476-484, 1993
 A:Title: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain stru
 A:Reference number: A47221; MUID:94010947
 A:Accession: A47221
 A:Molecule type: mRNA
 A:Residues: 1-337; 'T', 339-1029 <COR>
 A:Cross-references: GB:X63556
 R:Perleira, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Pangillan, T.; Bo
 Hum. Mol. Genet. 2, 961-968, 1993
 A:Title: Genomic organization of the sequence coding for fibrillin, the defective gen
 A:Reference number: I54355; MUID:93372860
 A:Accession: I54355
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA

A:Residues: 1332-3002 <PER>
A:Cross-references: GB:LI39923; NID:g306745; PIDN:AAB02036.1; PID:g306746
R:Mastlen, C.L.; Corson, G.W.; Maddox, B.K.; Glanville, R.W.; Sakai, L.Y.
Nature 352, 334-337, 1991
A>Title: Partial sequence of a candidate gene for the Marfan syndrome.
A:Reference number: S17064; MUID:91304568
A:Accession: S17064
A:Molecule type: mRNA
A:Residues: 1030-3002 <MAS>
A:Cross-references: EMBL:X63556
R:Dietz, H.C.; Valle, D.; Franccomano, C.A.; Kendzior, R.J.
Science 259, 680-683, 1993
A>Title: The skipping of constitutive exons in vivo induced by nonsense mutations.
A:Reference number: I59574; MUID:93157831
A:Accession: I59574
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 2217-2288, 'T', 2290-2325 <RES>
A:Cross-references: GB:S54426; NID:g264860; PIDN:AMB25244.1; PID:g264861
R:Lee, B.; Godfrey, M.; Vitale, E.; Horl, H.; Mettel, M.G.; Sarfarazi, M.; Tsipouras, P.
Nature 352, 330-334, 1991
A>Title: Linkage of Marfan syndrome and a phenotypically related disorder to two different
A:Reference number: S17062; MUID:91304567
A:Accession: S17062
A:Molecule type: mRNA
A:Residues: 'VLVAVDFILSYNKML', 944-1444 <LEB1>
A:Cross-references: EMBL:X62008; NID:g31398; PIDN:CAB56534.1; PID:g5924015
A:Accession: S62111
A:Molecule type: protein
A:Residues: 1166-1176, 'X', 1178-1180, 'D', 1182-1185 <LEE2>
R:Maddox, B.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W.
J. Biol. Chem. 264, 21381-21385, 1989
A>Title: Connective tissue microfibrils. Isolation and characterization of three large F
J. Biol. Chem. 264, 21381-21385, 1989
A:Reference number: A34198; MUID:90078246
A:Accession: A34198
A:Molecule type: protein
A:Residues: 565-575, 1890-1892, 'I', 1894-1900 <MA>
C:Comment: Fibrillin is a major component of elastin-associated microfibrils.
C:Genetics:
A:Gene: GDB:FBN1
A:Cross-references: GDB:I27115; OMIM:134797; OMIM:154700
A:Map position: 15q21.1-15q21.1
A:Introns: 2236/1; 2258/1; 2297/1
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein; N
F:13002/Product: fibrillin (5'-region exon A splice form) (fragment) #status predicted
F:133-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MTC>
F:1332-1367/Domain: EGF homology <EGF>
F:1457-1492/Domain: EGF homology <EGF>
F:2262-2295/Domain: EGF homology <EGF>

Query Match 17.1%; Score 113.5; DB 2; Length 3002;
Best Local Similarity 27.1%; Pred. No. 0.34;
Matches 36; Conservative 9; Mismatches 41; Indels 47; Gaps 7;

OY 23 CT-----ECYCH-----CTGGADCTCTDACTG--CGNCPPNAHTC----- 55
|| || || || || || || || || || || || || || || || || || || ||
Db 1343 CTNSEGSYECGQPEFALMPORRSTCIDDECDNPNICDDGGQCNTINPBYKCLCYDGMA 1402
OY 56 -TDSKNCYKAATCTGSTGCNTARTCTNSKCF-----EAKT-CTD-----ST 95
|| || || || || || || || || || || || || || || || || || || ||
Db 1403 SEDMKTCVDVNECDLPNPNICSGTCENTKGSGFICHDMGYSGGKKGKTCTDINECEIGAH 1462
OY 96 NCYKATACTNSTG 108
|| | ||| |
Db 1463 NCGKHAVCTNTAG 1475

RESULT 10
A35844
Xotch protein - African clawed frog
Species: Xenopus laevis (African clawed frog)
Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 13-Aug-1999

```
CcAccession: A35844
R:Coffman, C.; Harris, W.; Kintner, C.
Science 249, 1438-1441, 1990
A>Title: Notch, the xenopus homolog of Drosophila notch.
A:Reference number: A35844; MUID:90385285
A:Accession: A35844
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: mRNA
A:Residues: 1-2524 <CQF>
C:Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; EGF homo
C:Keywords: transmembrane protein
F:146-177/Domain: EGF homology <EGX1>
F:184-215/Domain: EGF homology <EGF1>
F:222-254/Domain: EGF homology <EGF>
E:456-487/Domain: EGF homology <EGX2>
E:757-788/Domain: EGF homology <EGF3>
F:1025-1056/Domain: EGF homology <EGX3>
F:1924-1956/Domain: ankryrin repeat homology <AN1>
F:1957-1989/Domain: ankryrin repeat homology <AN2>
F:1991-2023/Domain: ankryrin repeat homology <AN3>
F:2024-2056/Domain: ankryrin repeat homology <AN4>
F:2057-2089/Domain: ankryrin repeat homology <AN5>

Query Match          17.0%; Score 113; DB 2; Length 2524;
Best Local Similarity 34.3%; Pred.No. 0.34;

Matches   34; Conservative    5; Mismatches   42; Indels   18; Gaps      6;

Qy   23 CRECY----CHCT---GGADCTSCIDACTGCGNCNPAHTCDSKKCVAAAC-----TG 69
      ||:|||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||
Db   960 CPDVCNVSTYTCIQPFGSGIHCESNTPDCTE--SSCFNGGTGIDG--INFTQCQCPPGFTG 1015
      |::|:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||

Qy   70 STCKTARTCTNSKDCFEAKTCTDSTNCYKATACYNSTG 108
      |::|:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||
Db   1016 SYCQHDIRMC-DKRPCPLNGGTQGDVGYTKYKCPCPGYTG 1053
      |::|:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||:-||

RESULT 11
S18946
      ultra high-sulfur keratin 1 - human
N:Alternate names: UHS keratin; ultra high-sulfur matrix protein
C:Species: Homo sapiens (man)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: S18946; B36686
R:Draebent, B.; Doebecke, D.
submitted to the EMBL Data Library, December 1991
A:Description: Nucleotide sequence of a Human high-sulphur keratin cDNA.
A:Reference number: S18946
A:Accession: S18946
A:Molecule type: mRNA
A:Residues: 1-169 <DRA>
A:Cross-references: EMBL:X63755; NID:g32471; PIDN:CAA45283.1; PID:g32472
R:MacKinnon, P.J.; Powell, B.C.; Rogers, G.E.
J. Cell Biol. 111, 2587-2600, 1990
A>Title: Structure and expression of genes for a class of cysteine-rich proteins of t
A:Reference number: A36686; MUID:91115951
A:Accession: B36686
A:Molecule type: DNA
A:Residues: 1-59, 'Y', 41-169 <MAC>
A:Cross-references: GB:X5293; NID:g34078; PIDN:CAA39005.1; PID:g34079
C:Gene: GDB:KRNI
A:Genetics:
A:Cross-references: GDB:125257; OMIM:148021
A:Map position: 11q13-11q13
C:Superfamily: ultra-high-sulfur keratin
C:Keywords: hair; tandem repeat
F:7-15/Region: Ser-rich nonapeptide repeat
F:59-68/Region: Gly-rich decapeptide repeat
F:69-78/Region: Gly-rich decapeptide repeat
F:79-88/Region: Cys-rich decapeptide repeat
F:89-97/Region: Ser-rich nonapeptide repeat
F:98-107/Region: Cys-rich decapeptide repeat
F:108-117/Region: Cys-rich decapeptide repeat
F:118-126/Region: Ser-rich nonapeptide repeat
```

F:127-136/Region: Cys-rich decapeptide repeat
 F:137-145/Region: Ser-rich nonapeptide repeat
 F:146-155/Region: Cys-rich decapeptide repeat
 F:156-165/Region: Cys-rich decapeptide repeat

Query Match 16.9%; Score 112.5; DB 1; Length 169;
 Best Local Similarity 26.5%; Pred. No. 0.081;
 Matches 27; Conservative 14; Mismatches 44; Indels 17; Gaps 4;
 Oy 19 VMCLTECHCHCTGADCTCTDCTACTGGGNCNPAHCTCDNSNCKATCTGSG-----70
 Db 45 VCCCVAPSCSSCKRGCGSGGSGSGC-GCGSCCCKPCCSSGCGSSCCGCCSCCK 103
 Oy 71 ---FKNTARTCTNSKDCFEAKTCTDSTNCKATCTNSTGC 109
 Db 104 PYCSGCSCKPCCSSG--RSSGCCSSC---KPCSSSGC 140

RESULT 12
 T26972
 hypothetical protein Y47H9C.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 C:Accession: T26972
 R:Harris, B.
 submitted to the EMBL Data Library, October 1998
 A:Reference number: Z20293
 A:Accession: T26972
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1111 <WTL>
 A:Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4
 A:Experimental source: clone Y47H9C
 C:Genetics:
 A:Gene: CESP:Y47H9C.4
 A:Map position: 1
 A:Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

Query Match 16.9%; Score 112; DB 2; Length 1111;
 Best Local Similarity 30.4%; Pred. No. 0.25;
 Matches 31; Conservative 4; Mismatches 43; Indels 24; Gaps 6;
 Oy 28 CHCTGADCTG-----CTDCTGCGNCPNAHCTDSK---NCVKAATCTGTCNTART 78
 Db 164 CDCENGANDPBLCTCTSGFG-ERCEK--PCPDNKGNCVKSCCGCGCKNEKG 220
 Oy 79 CTNS-----KDCFEAK--TCCTDSTNCKATCTNSTG 108
 Db 221 CVCSIDGGEFCINCKCEGKFGAECKFCNCNGATCNDTNG 262

RESULT 13
 C42125
 trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)
 N:Alternate names: CRP72
 C:Species: Giardia lamblia
 C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998
 C:Accession: C42125
 R:Adam, R.D.; Yang, Y.M.; Nash, T.E.
 M.Ol. Cell. Biol. 12, 1194-1201, 1992
 A>Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170 gene
 A:Reference number: M42125; MUID:92186850
 A:Accession: C42125
 A:Molecule type: DNA
 A:Residues: 1-677 <ADA>
 A:Cross-references: GB:M83934; NID:g159123
 A:Experimental source: trophozoites
 A>Note: sequence extracted from NCBI backbone (NCBI:88443, NCBI:P:88444); this ORF is not
 C:Keywords: surface antigen

Query Match 16.7%; Score 111; DB 2; Length 677;
 Best Local Similarity 29.2%; Pred. No. 0.23;

Matches 35; Conservative 13; Mismatches 36; Indels 36; Gaps 8;
 Oy 22 LCTECHCHCTG-----GADCTCT-----DAGCGNCPNAHCTDSKNKYAA 65
 Db 104 LCTECKT-ANGLEKPAAPPEKSECTILSDINGAGYGVANCAO--CTKSDSNKGA 159
 Oy 66 TCT-----GSTKCN-TARTC-TNSKDCFEAKTCTDSTNCKATCTNSTGCPGH 112
 Db 160 TCTACGAGYKRPQASCKDGTCTCTETSAQC---TSCPEGYLKGASCVNNNGCTGN 216

RESULT 14
 T16840
 hypothetical protein T10E10.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16840
 R:Geisel, C.
 submitted to the EMBL Data Library, October 1995
 A:Description: The sequence of C. elegans cosmid T10E10.
 A:Reference number: Z18588
 A:Accession: T16840
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1101 <GEI>
 A:Cross-references: EMBL:U39644; NID:g1049339; PID:g1049343; PIDN:AAA80360.1; CESP:T1
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:T10E10.4
 A:Introns: 93/2; 152/2; 191/3; 209/2; 283/3; 303/1; 399/3; 421/1; 440/1; 465/1; 547/3

Query Match 16.7%; Score 111; DB 2; Length 1101;
 Best Local Similarity 25.4%; Pred. No. 0.3;
 Matches 30; Conservative 4; Mismatches 50; Indels 34; Gaps 5;
 Oy 23 CTETCHCTGCGADC-----TCTDCTGCGNCPNAHCTDSKNC-----61
 Db 661 CGRIGVDCNNGACCPPTCPNINIASQRCSGGCTNC--CPVGQTCMNGGCCDILPSPSGG 718
 Oy 62 VKAATCTGSKCMNARTCTNSKDC-----FEAKTCTDSTNCKATCTNSTGCP 110
 Db 719 FAISMCTG--KCGTGEVCNVSGCCSLPCPSGLMSVORCAMGIGCPPNGCENGVCPP 774

RESULT 15
 A55567
 fibrillin I - bovine
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 11-Jan-2000
 C:Accession: A55567
 R:Tilstra, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.
 Genomics 23, 480-485, 1994
 A>Title: Sequence number of the coding region of the bovine fibrillin cDNA and localization
 A:Reference number: A55567; MUID:95137597
 A:Accession: A55567
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2671 <TTI>
 A:Cross-references: GB:I28748; NID:g508427; PIDN:AAA74122.1; PID:g508428
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:1201-1236/Domain: EGF homology <EGF>

Query Match 16.6%; Score 110.5; DB 2; Length 2871;
 Best Local Similarity 26.3%; Pred. No. 0.55;
 Matches 35; Conservative 10; Mismatches 41; Indels 47; Gaps 7;
 Oy 23 CT-----ECYCH-----CTGADCTCTDCTG--CGNCPNAHCTC-----55
 Db 1212 CTNSEGSEYSCQPGFALMPDRSCDIDECEDNPNICDGGCTNIPGEYRCICYGFWA 1271
 Oy 56 -TDSKNKYKAATCTGSKCNTARTCTNSKDCF-----EAT-CTD-----ST 95
 Db 1272 SEDMKTCVDVNECDLNPNICLSGTCEWTKGSFICHCMDMGYSGRKKGTYGCTDINECEIGH 1331

OY 96 NCYKATCTNSTG 108
||: ||: |
Db 1332 NCDRHAVCTNTAG 1344

Search completed: October 21, 2002, 16:40:12
Job time : 25.2353 secs

| | | | | |
|----------|--------|----------|--------|-----------------------------|
| FT | REPEAT | 126 | 149 | A-2. |
| FT | REPEAT | 150 | 173 | B-1. |
| FT | REPEAT | 174 | 197 | A-3. |
| FT | REPEAT | 198 | 221 | A-4. |
| FT | REPEAT | 222 | 245 | B-2. |
| FT | REPEAT | 246 | 269 | A-5. |
| FT | REPEAT | 270 | 293 | A-6. |
| FT | REPEAT | 294 | 317 | B-3. |
| FT | REPEAT | 318 | 341 | A-7. |
| FT | REPEAT | 342 | 365 | A-8. |
| FT | REPEAT | 366 | 389 | B-4. |
| FT | REPEAT | 390 | 413 | A-9. |
| FT | REPEAT | 414 | 437 | A-10. |
| FT | REPEAT | 438 | 461 | B-5. |
| FT | REPEAT | 462 | 485 | A-11. |
| FT | REPEAT | 486 | 509 | A-12. |
| FT | REPEAT | 510 | 533 | B-6. |
| FT | REPEAT | 534 | 557 | A-13. |
| FT | REPEAT | 558 | 581 | X-4. |
| FT | REPEAT | 582 | 605 | A-14. |
| FT | REPEAT | 606 | 629 | A-15. |
| FT | REPEAT | 630 | 653 | A-16. |
| FT | REPEAT | 654 | 677 | B-7. |
| FT | REPEAT | 678 | 701 | A-17. |
| FT | REPEAT | 702 | 725 | A-18. |
| FT | REPEAT | 726 | 749 | B-8. |
| FT | REPEAT | 750 | 773 | A-19. |
| FT | REPEAT | 774 | 797 | A-20. |
| FT | REPEAT | 798 | 821 | B-9. |
| FT | REPEAT | 822 | 845 | A-21. |
| FT | REPEAT | 846 | 869 | A-22. |
| FT | REPEAT | 870 | 893 | A-23. |
| FT | REPEAT | 894 | 917 | B-10. |
| FT | REPEAT | 918 | 941 | A-24. |
| FT | REPEAT | 942 | 964 | A-25. |
| FT | REPEAT | 965 | 988 | A-26. |
| FT | REPEAT | 989 | 1012 | A-27. |
| SEQUENCE | | 1046 AA: | 108823 | MMV: F9DC04FD7D85818 CRC64; |

```

Query Match: 19.4%; Score 128.5; DB 1, Length 1046;
Best Local Similarity 35.9%; Pred. No. 0.00024;
Matches 33; Conservative 9; Mismatches 37; Indels 13; Gaps 5.

QY 23 CTCTCCHCTGTGADCT--SCTD--ACTGGGNCNPHATCTDSKNCKVA-ATCTGSTKCTAR 77
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 800 CTEDKCTQNGNGVYHFTIRCDLINST-----ADSCNSISGCVHFTPIINDDNKKC-TAD 851

QY 78 TCINSKDCFEAKTCTDSTNCYKATACTNSTGC 109
      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 852 SCNSNSTGCCHTPISCDNNPCTVDCSNSTGC 883

RESULT 2
LRP2_RAT
ID LRP2_RAT STANDARD: PRT. 4660 AA.
AC P98158;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Low-density lipoprotein receptor-related protein 2 precursor (Megalin)
DE (Glycoprotein 330).
LN LRP2.
GN Rattus norvegicus (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;
RX MEDLINE=95024033; PubMed=7937880;
RA Saito A., Pietromonaco S., Loo A.K.C., Faruqhar M.G.;
RT 'Complete cloning and sequencing of rat gp330/megalin,' a
RT distinctive member of the low density lipoprotein receptor gene

```

| | | |
|----|--|---|
| RT | | family."; |
| RL | | Proc. Natl. Acad. Sci. U.S.A. 91:9725-9729(1994). |
| RN | [2] | |
| RP | | FUNCTION. |
| RX | MEDLINE=95386696; PubMed=7544804; | |
| RA | Moestlup S.K., Cui S., Vorum H., Bregengaard C., Bjorn S.E., | |
| RA | Norris K., Gilemann J., Christensen E.I.;" | |
| RT | "Evidence that epithelial glycoprotein 330/megalin mediates uptake of | |
| RT | polybasic invst." | |
| RL | J. Clin. Invest. 96:1404-1413(1995). | |
| RN | [3] | |
| RP | TISSUE SPECIFICITY. | |
| RA | MEDLINE=94172242; PubMed=7510321; | |
| RA | Zheng G., Bachinsky D.R., Stamenkovic I., Strickland D.K., Brown D., | |
| RA | Andres G., McCluskey R.T.;" | |
| RT | "Organ distribution in rats of two members of the low-density | |
| RT | lipoprotein receptor gene family, gp330 and Lrp/alpa ZMR, and the | |
| RL | receptor-associated protein (RAP)."; | |
| RL | J. Biochem. Cytochem. 42:531-542(1994). | |
| CC | - FUNCTION: BINDS PLASMINOGEN, EXTRACELLULAR MATRIX COMPONENTS, | |
| CC | PLASMINOGEN ACTIVATOR-PLASMINOGEN ACTIVATOR INHIBITOR TYPE I, | |
| CC | COMPLEX, APOLOPROTEIN E-ENRICHED BETA-VLDL, LIPOPROTEIN LIPASE, | |
| CC | LACTOFERRIN, CLUSTERIN AND CALCIUM. | |
| CC | - FUNCTION: RECEPTOR-MEDIATED UPTAKE OF POLYBASIC DRUGS SUCH AS | |
| CC | APROTININ, AMINOGLYCOSIDES AND POLYMYXIN B. | |
| CC | - SUBUNIT: FORMS A MULTIMERIC COMPLEX TOGETHER WITH A RECEPTOR- | |
| CC | ASSOCIATED PROTEIN (RAP). | |
| CC | - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXPRESSED IN | |
| CC | CLATHRIN-COATED PITS; A SOLUBLE FORM IS POSSIBLY DERIVED BY | |
| CC | CLEAVAGE AT THE CELL SURFACE. | |
| CC | - TISSUE SPECIFICITY: EPITHELIAL CELLS OF KIDNEY GLOMERULUS AND | |
| CC | PROXIMAL TUBULE, LUNG, EPIDIDYMIS, YOLK SAC, AMONG OTHERS. | |
| CC | - SIMILARITY: CONTAINS 36 LDL-RECEPTOR CLASS A DOMAINS. | |
| CC | - SIMILARITY: CONTAINS 37 LDL-RECEPTOR CLASS B DOMAINS. | |
| CC | - SIMILARITY: CONTAINS 17 EGF-LIKE DOMAINS. | |
| CC | ----- | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL Outstation - | |
| CC | the European Bioinformatics Institute. There are no restrictions on ways | |
| CC | use by non-profit institutions as long as its content is in no way com- | |
| CC | modified and this statement is not removed. Usage by and for commercial | |
| CC | entities requires a license agreement (See http://www.isb-slb.ch/announce/ | |
| CC | or send an email to license@isb-slb.ch). | |
| CC | ----- | |
| CC | EMBL; L34049; AAAS1369.1; . | |
| DR | HSSP; Q07954; ICR8. | |
| DR | GlycosultedB; P98158; . | |
| DR | InterPro: IPRO00152; Asx_hydroxyl. | |
| DR | InterPro: IPRO00561; EGF-like. | |
| DR | InterPro: IPRO01881; EGF_Ca. | |
| DR | InterPro: IPRO02172; LDL_recept_A. | |
| DR | InterPro: IPRO00033; LDL_receptor_rep. | |
| DR | Pfam; PF00057; ldl_recept_a; 36. | |
| DR | Pfam; PF00058; ldl_recept_b; 33. | |
| DR | PRINTS; PR00261; LDLRECEPTOR. | |
| DR | SMART; SM00179; EGF_CA_3. | |
| DR | SMART; SM00001; EGF_like; 15. | |
| DR | SMART; SM00192; LDla; 36. | |
| DR | SMART; SM00135; LY; 35. | |
| DR | PROSITE; PS00010; ASX_HYDROXYL; 4. | |
| DR | PROSITE; PS00022; EGF_1; 1. | |
| DR | PROSITE; PS01186; EGF_2; 8. | |
| DR | PROSITE; PS01187; EGF_CA; 3. | |
| DR | PROSITE; PS01209; LDLRA_1; 31. | |
| DR | PROSITE; PS50068; LDLRA_2; 36. | |
| KW | Glycoprotein; Repeat; Endocytosis; Coated pits; Transmembrane; | |
| KW | Receptor; EGF-like domain; SH3-binding; Signal. | |
| FT | SIGNAL | |
| FT | 1 25 | |
| FT | CHAIN | |
| FT | 26 4660 | |
| FT | | LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED |
| FT | | PROTEIN 2. |
| FT | DNAIN | |
| FT | 26 4425 | |
| FT | TRANSMEM | |
| FT | 4426 4446 | |
| FT | 4447 4660 | |
| FT | | EXTRACELLULAR (POTENTIAL). |
| FT | | POTENTIAL. |
| FT | | CYTOPASMIC (POTENTIAL). |

```
FT DOMAIN 26 64 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 65 105 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 106 144 LDL-RECEPTOR CLASS A 3.
FT DOMAIN 145 181 LDL-RECEPTOR CLASS A 4.
FT DOMAIN 181 219 LDL-RECEPTOR CLASS A 5.
FT DOMAIN 220 258 LDL-RECEPTOR CLASS A 6.
FT DOMAIN 263 307 LDL-RECEPTOR CLASS A 7.
FT DOMAIN 308 346 EGF-LIKE 1.
FT DOMAIN 347 385 EGF-LIKE 2.
FT DOMAIN 435 476 LDL-RECEPTOR CLASS B 1.
FT DOMAIN 478 519 LDL-RECEPTOR CLASS B 2.
FT DOMAIN 521 566 LDL-RECEPTOR CLASS B 3.
FT DOMAIN 568 611 LDL-RECEPTOR CLASS B 4.
FT DOMAIN 612 652 LDL-RECEPTOR CLASS B 5.
FT DOMAIN 658 704 EGF-LIKE 3.
FT DOMAIN 752 793 LDL-RECEPTOR CLASS B 6.
FT DOMAIN 795 835 LDL-RECEPTOR CLASS B 7.
FT DOMAIN 837 879 LDL-RECEPTOR CLASS B 8.
FT DOMAIN 881 923 LDL-RECEPTOR CLASS B 9.
FT DOMAIN 969 1013 EGF-LIKE 4.
FT DOMAIN 1023 1061 LDL-RECEPTOR CLASS A 8.
FT DOMAIN 1064 1103 LDL-RECEPTOR CLASS A 9.
FT DOMAIN 1108 1146 LDL-RECEPTOR CLASS A 10.
FT DOMAIN 1148 1186 LDL-RECEPTOR CLASS A 11.
FT DOMAIN 1186 1225 LDL-RECEPTOR CLASS A 12.
FT DOMAIN 1229 1269 LDL-RECEPTOR CLASS A 13.
FT DOMAIN 1270 1308 LDL-RECEPTOR CLASS A 14.
FT DOMAIN 1311 1351 LDL-RECEPTOR CLASS A 15.
FT DOMAIN 1350 1390 EGF-LIKE 5.
FT DOMAIN 1391 1430 EGF-LIKE 6.
FT DOMAIN 1430 1479 LDL-RECEPTOR CLASS B 10.
FT DOMAIN 1479 1520 LDL-RECEPTOR CLASS B 11.
FT DOMAIN 1522 1563 LDL-RECEPTOR CLASS B 12.
FT DOMAIN 1567 1609 LDL-RECEPTOR CLASS B 13.
FT DOMAIN 1611 1654 LDL-RECEPTOR CLASS B 14.
FT DOMAIN 1656 1696 LDL-RECEPTOR CLASS B 15.
FT DOMAIN 1701 1742 EGF-LIKE 7.
FT DOMAIN 1742 1782 LDL-RECEPTOR CLASS B 16.
FT DOMAIN 1832 1882 LDL-RECEPTOR CLASS B 17.
FT DOMAIN 1884 1930 LDL-RECEPTOR CLASS B 18.
FT DOMAIN 1932 1972 LDL-RECEPTOR CLASS B 19.
FT DOMAIN 1973 2013 EGF-LIKE 8.
FT DOMAIN 2019 2060 LDL-RECEPTOR CLASS B 20.
FT DOMAIN 2108 2156 LDL-RECEPTOR CLASS B 21.
FT DOMAIN 2158 2201 LDL-RECEPTOR CLASS B 22.
FT DOMAIN 2203 2245 LDL-RECEPTOR CLASS B 23.
FT DOMAIN 2247 2289 LDL-RECEPTOR CLASS B 24.
FT DOMAIN 2291 2332 EGF-LIKE 9.
FT DOMAIN 2332 2384 LDL-RECEPTOR CLASS B 25.
FT DOMAIN 2432 2477 LDL-RECEPTOR CLASS B 26.
FT DOMAIN 2479 2518 LDL-RECEPTOR CLASS B 27.
FT DOMAIN 2520 2562 LDL-RECEPTOR CLASS B 28.
FT DOMAIN 2564 2604 LDL-RECEPTOR CLASS B 29.
FT DOMAIN 2605 2647 EGF-LIKE 10.
FT DOMAIN 2652 2694 LDL-RECEPTOR CLASS A 16.
FT DOMAIN 2699 2739 LDL-RECEPTOR CLASS A 17.
FT DOMAIN 2740 2778 LDL-RECEPTOR CLASS A 18.
FT DOMAIN 2779 2820 LDL-RECEPTOR CLASS A 19.
FT DOMAIN 2821 2862 LDL-RECEPTOR CLASS A 20.
FT DOMAIN 2863 2903 LDL-RECEPTOR CLASS A 21.
FT DOMAIN 2906 2947 LDL-RECEPTOR CLASS A 22.
FT DOMAIN 2948 2992 LDL-RECEPTOR CLASS A 23.
FT DOMAIN 3031 3072 LDL-RECEPTOR CLASS A 24.
FT DOMAIN 3072 3112 LDL-RECEPTOR CLASS A 25.
FT DOMAIN 3113 3153 EGF-LIKE 11.
FT DOMAIN 3153 3194 EGF-LIKE 12.
FT DOMAIN 3194 3241 LDL-RECEPTOR CLASS B 30.
FT DOMAIN 3241 3282 LDL-RECEPTOR CLASS B 31.
FT DOMAIN 3284 3333 LDL-RECEPTOR CLASS B 32.
FT DOMAIN 3335 3377 LDL-RECEPTOR CLASS B 33.
FT DOMAIN 3379 3420 LDL-RECEPTOR CLASS B 34.
FT DOMAIN 3421 3461 EGF-LIKE 13.
FT DOMAIN 3467 3511 LDL-RECEPTOR CLASS A 26.
FT DOMAIN 3512 3552
```

| | | | | | | |
|----|----------|------|------|----------------|-----------------|-------------|
| FT | DOMAIN | 566 | 601 | EGF-LIKE 15, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 603 | 639 | EGF-LIKE 16, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 641 | 676 | EGF-LIKE 17, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 678 | 714 | EGF-LIKE 18, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 716 | 751 | EGF-LIKE 19, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 753 | 789 | EGF-LIKE 20, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 791 | 827 | EGF-LIKE 21, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 829 | 867 | EGF-LIKE 22, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 869 | 905 | EGF-LIKE 23, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 907 | 943 | EGF-LIKE 24, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 945 | 981 | EGF-LIKE 25, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 983 | 1019 | EGF-LIKE 26, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1021 | 1057 | EGF-LIKE 27, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1059 | 1095 | EGF-LIKE 28, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1097 | 1143 | EGF-LIKE 29, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1145 | 1181 | EGF-LIKE 30, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1183 | 1219 | EGF-LIKE 31, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1221 | 1265 | EGF-LIKE 32, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1267 | 1305 | EGF-LIKE 33, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1307 | 1346 | EGF-LIKE 34, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1348 | 1384 | EGF-LIKE 35, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1387 | 1426 | EGF-LIKE 36, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DOMAIN | 1449 | 1462 | CYS-RICH, | CALCIUM-BINDING | (POTENTIAL) |
| FT | REPEAT | 1445 | 1480 | LIN/NOTCH 1, | CALCIUM-BINDING | (POTENTIAL) |
| FT | REPEAT | 1481 | 1522 | LIN/NOTCH 2, | CALCIUM-BINDING | (POTENTIAL) |
| FT | REPEAT | 1523 | 1562 | LIN/NOTCH 3, | CALCIUM-BINDING | (POTENTIAL) |
| FT | REPEAT | 1917 | 1947 | ANK 1, | CALCIUM-BINDING | (POTENTIAL) |
| FT | REPEAT | 1949 | 1979 | ANK 2, | CALCIUM-BINDING | (POTENTIAL) |
| FT | REPEAT | 1983 | 2012 | ANK 3, | CALCIUM-BINDING | (POTENTIAL) |
| FT | REPEAT | 2016 | 2045 | ANK 4, | CALCIUM-BINDING | (POTENTIAL) |
| FT | REPEAT | 2049 | 2078 | ANK 5, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 24 | 37 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 31 | 46 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 63 | 74 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 89 | 87 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 88 | 98 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 106 | 117 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 111 | 127 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 129 | 138 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 144 | 155 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 149 | 164 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 166 | 175 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 182 | 195 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 189 | 204 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 206 | 215 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 222 | 233 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 227 | 243 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 245 | 254 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 261 | 272 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 266 | 281 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 283 | 292 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 299 | 312 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 306 | 321 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 323 | 332 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 339 | 350 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 344 | 359 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 361 | 370 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 376 | 387 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 381 | 398 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 400 | 409 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 416 | 429 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 423 | 438 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 440 | 449 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | 456 | 467 | BY SIMILARITY, | CALCIUM-BINDING | (POTENTIAL) |
| FT | DISULFID | | | | | |

```

FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 987 998 BY SIMILARITY.
FT DISULFID 992 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.

Query Match 17.7% Score 117.5; DB 1; Length 2531;
Best Local Similarity 32.1%; Pred. No. 0.00411;
Matches 35; Conservative 7; Mismatches 46; Indels 21; Gaps 6;

QY 15 IAVIWC-----LCTE-----CYCHCTGAGDCSTCDCTGCG--GNCPRNHTCTD--SKNC 61
DB 1110 IIVTLICQGGGLCVDEBGRKHCHQAGTGYCDEVEBCSPNQGNATCTDYIGGRSC 1169
QY 62 VAAATCTGSKCTNARTCTNSKDCFEAKTCTDSTNCTYACTACTNCTGCP 110
DB 1170 KCVAGYHGSGNCSEINECL-SQPCONGGTCTDITNSYKCS-----CP 1210

RESULT 4
NTCL_RAT STANDARD: PRT; 2531 AA.
AC Q07008;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuronic locus notch homolog protein 1 precursor.
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Schwann cell;
RX MEDLINE=92111383; PubMed=1764995;
RA Weimaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian
RL development.";
RN (2)
RP REVISIONS TO 1652-1653.

```

```

RA Weimaster G.;
RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC OF TISSUES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC ADULT.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57405; CAA00667.1; -.
DR HSSP; P00740; IIXA.
DR InterPro; IPRO02110; ANK.
DR InterPro; IPRO00152; Asx_hydroxyl.
DR InterPro; IPRO00561; EGF-like.
DR InterPro; IPRO00742; EGF-2.
DR InterPro; IPRO01881; EGF_Ca.
DR InterPro; IPRO01438; EGF_II.
DR InterPro; IPRO02049; Lamln_EGF.
DR InterPro; IPRO00800; Notch.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBLDOD.
DR PRINTS; PR00011; EGFFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 25.
DR SMART; SM00001; EGF_Like; 10.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 35.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 2531
FT DOMAIN 19 1723
FT TRANSMEM 1724 1746
FT DOMAIN 1747 2531
FT DOMAIN 20 58
FT DOMAIN 59 99
FT DOMAIN 102 139
FT DOMAIN 140 176
FT DOMAIN 178 216
FT DOMAIN 218 255
FT DOMAIN 257 293
FT DOMAIN 295 333
FT DOMAIN 335 371
FT DOMAIN 372 410
FT DOMAIN 412 488
FT DOMAIN 452 480
FT DOMAIN 490 526
FT DOMAIN 528 564
FT DOMAIN 566 601
FT DOMAIN 603 639
FT DOMAIN 641 676
FT DOMAIN 678 714
FT DOMAIN 716 751

```

```

FT DOMAIN 753 789 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 829 867 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 869 905 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 907 943 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 945 981 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 983 1019 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1021 1057 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1059 1095 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1097 1143 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1145 1181 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1183 1219 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1221 1265 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1267 1305 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1307 1346 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1348 1384 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1387 1426 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1449 1462 CYS-RICH.
FT REPEAT 1917 1946 ANK 1.
FT REPEAT 1950 1980 ANK 2.
FT REPEAT 1984 2013 ANK 3.
FT REPEAT 2017 2046 ANK 4.
FT REPEAT 2050 2079 ANK 5.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 589 591 BY SIMILARITY.
FT DISULFID 600 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.

```

```

FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 807 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 844 855 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 967 998 BY SIMILARITY.
FT DISULFID 992 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1068 1083 BY SIMILARITY.
FT DISULFID 1085 1094 BY SIMILARITY.
FT DISULFID 1101 1122 BY SIMILARITY.
FT DISULFID 1116 1131 BY SIMILARITY.

```

Query Match 17.7%; Score 117.5; DB 1; Length 2531;
 Best Local Similarity 32.1%; Pred. No. 0.0041;
 Matches 35; Conservative 7; Mismatches 46; Indels 21; Gaps 6;

```

QY 15 IAVIWM-----LCTE---CYCHCTGADCTCTACTGCG--GNCPRMHTCTD--SKNC 61
DB 1110 IDVTLLCGHGGLCVDEEDKHCHCQAGTIGSYCEDEVCSPNFCQNCATCTDIYIGRSC 1169
QY 62 VKAATCTGTFKCNRTARTCTNSKDCFEAKTCTDSTNCYKATACNSTGCP 110
DB 1170 KCVAGYHGSGNCSEETNECL-SQPCQNGGTCTDITWTYCS-----CP 1210

```

RESULT 5

```

KRUB_HUMAN
ID KRUB_HUMAN STANDARD; PRT; 194 AA.
AC 075690;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, ultra high-sulfur matrix protein B (UHS keratin B) (UHS
  DE Kerb).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148005; Pubmed=10023043;
RA Perez C., Aurio J., Garst C., Bernard B.A., Egly J.-M.;
RT "Genomic organization and promoter characterization of two human UHS
  RT keratin genes."
RL Gene 227:137-148(1999).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).

```



```

FT REPEAT 1928 1957 ANK 1.
FT REPEAT 1961 1991 ANK 2.
FT REPEAT 1995 2024 ANK 3.
FT REPEAT 2028 2057 ANK 4.
FT REPEAT 2061 2090 ANK 5.
FT DOMAIN 1576 1579 POLY-VAL.
FT DOMAIN 1662 1665 POLY-ARG.
FT DOMAIN 1729 1732 POLY-PRO.
FT DOMAIN 1741 1744 POLY-ALA.
FT DOMAIN 1902 1905 POLY-GLU.
FT DOMAIN 2260 2263 POLY-GLY.
FT DOMAIN 2404 2407 POLY-GLN.
FT DOMAIN 2411 2418 POLY-PRO.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.

```

```

FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 867 BY SIMILARITY.
FT DISULFID 874 885 BY SIMILARITY.
FT DISULFID 879 894 BY SIMILARITY.
FT DISULFID 896 905 BY SIMILARITY.
FT DISULFID 912 923 BY SIMILARITY.
FT DISULFID 917 932 BY SIMILARITY.
FT DISULFID 934 943 BY SIMILARITY.
FT DISULFID 988 999 BY SIMILARITY.
FT DISULFID 993 1008 BY SIMILARITY.
FT DISULFID 1010 1019 BY SIMILARITY.
FT DISULFID 1026 1037 BY SIMILARITY.
FT DISULFID 1031 1046 BY SIMILARITY.
FT DISULFID 1048 1057 BY SIMILARITY.
FT DISULFID 1064 1075 BY SIMILARITY.

```

Query Match 17.2%; Score 114; DB 1; Length 2444;

Best Local Similarity 31.6%; Pred.No.0.0081; Matches 31; Conservative 6; Mismatches 45; Indels 16; Gaps 6;

```

QY 23 CTE----CYCHCTGADCTCTDCTGCGN--CPNAHCTDSKNCVKATCT-----GST 71
Db 923 CTGGINTPAFCDLPGRGIFCEEDINECASDPCRNANCND---CYDSTYTCIPAFSGI 979
QY 72 KC-NTARTCTNSKDCFEARTCTDSTNCYKATACTNSTG 108
Db 980 HCENNTPDCTES--SCFNGGTCVDGINSFTCLCPGFTG 1016

```

RESULT 7

```

ID FBNI_HUMAN STANDARD; PRT; 2871 AA.
AC P35555;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibrillin 1 precursor.
GN FBNI OR FBN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA Pangillan T., Bonadio J.;
RT "Genomic organization of the sequence coding for fibrillin, the
RT defective gene product in Marfan syndrome.";
RL Hum. Mol. Genet. 2:961-968(1993).
RN [2]
RP SEQUENCE OF 1-932 FROM N.A.
RC TISSUE=Placenta, and fibroblast;
RA MEDLINE=94010947; PubMed=7691719;
RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;
RT "Fibrillin binds calcium and is coded by cDNAs that reveal a
RT multidomain structure and alternatively spliced exons at the 5'
RT end.";
RL Genomics 17:476-484(1993).
RN [3]
RP SEQUENCE OF 899-2871 FROM N.A.
RA MEDLINE=91304568; PubMed=1852207;
RA Maslen C.L., Corson G.M., Maddox B.K., Glanville R.W., Sakai L.Y.;
RT "Partial sequence of a candidate gene for the Marfan syndrome.";
RL Nature 352:334-337(1991).
RN [4]
RP SEQUENCE OF 813-1313 FROM N.A.
RA MEDLINE=91304567; PubMed=1852206;

```


RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
 RA Tsiipouras P., Ramirez F., Hollister D.W.;
 RA "Linkage of Marfan syndrome and a phenotypically related disorder to
 RT two different fibrillin genes."; *Nature* 352:330-334(1991).
 RL [15]
 RN CHARACTERIZATION.
 RP MEDLINE=91317849; PubMed=1860873;
 RA Sakai L.Y., Keene D.R., Glanville R.W., Bachinger H.P.;
 RA "Purification and partial characterization of fibrillin, a cysteine-
 RT rich structural component of connective tissue microfibrils."; *J.
 RL Biol. Chem.* 266:14763-14770(1991).
 RN [16]
 RP STRUCTURE BY NMR OF 2054-2125.
 RX MEDLINE=98031893; PubMed=9362480;
 RA Yuan X., Downing A.K., Knott V., Handford P.A.;
 RA "Solution structure of the transforming growth factor beta-binding
 RT protein-like module, a domain associated with matrix fibrils."; *EMBO J.*
 RL 16:6659-6666(1997).
 RN [17]
 RP STRUCTURE BY NMR OF 2124-2205.
 RX MEDLINE=96144829; PubMed=8556869;
 RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;
 RA "Calcium binding properties of an epidermal growth factor-like domain
 RT pair from human fibrillin-1."; *J. Mol. Biol.* 255:22-27(1996).
 RN [18]
 RP STRUCTURE BY NMR OF 2124-2205.
 RX MEDLINE=9622301; PubMed=8653794;
 RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,
 RA Handford P.A.;
 RA "Solution structure of a pair of calcium-binding epidermal growth
 RT factor-like domains: implications for the Marfan syndrome and other
 RL genetic disorders."; *Cell* 85:597-605(1996).
 RN [19]
 RP REVIEW ON MFS VARIANTS.
 RX MEDLINE=96174615; PubMed=8594563;
 RA Colod-C, Beroud C., Soussi T., Junien C., Boileau C.;
 RA "Software and database for the analysis of mutations in the human
 RT FBN1 gene."; *Nucleic Acids Res.* 24:137-141(1996).
 RL [10]
 RP REVIEW ON MFS VARIANTS.
 RX MEDLINE=97169383; PubMed=9016526;
 RA Colod-Beroud G., Beroud C., Ades L., Black C., Boxer M., Brock D.J.,
 RA Godfrey M., Hayward C., Karttunen L., Milewicz D., Peltonen L.,
 RA Richards R.I., Wang W., Junien C., Boileau C.;
 RA "Marfan Database (second edition): software and database for the
 RT analysis of mutations in the human FBN1 gene."; *Nucleic Acids Res.*
 RL 25:147-150(1997).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=98062175; PubMed=9401003;
 RA Hayward C., Brock D.J.H.;
 RA "Fibrillin-1 mutations in Marfan syndrome and other type-1
 RT fibrillinopathies."; *Hum. Mutat.* 10:415-423(1997).
 RL [12]
 RP VARIANT MFS PRO-1137.
 RX MEDLINE=91304569; PubMed=1852208;
 RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,
 RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,
 RA Curristin S.M., Stetten G., Meyers D.A., Francomano C.A.;
 RA "Marfan syndrome caused by a recurrent de novo missense mutation in
 RT the fibrillin gene."; *Nature* 352:337-339(1991).
 RL [13]
 RP VARIANT MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.
 RX MEDLINE=93250834; PubMed=1301946;
 RA Dietz H.C., Saraiya J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;
 RA "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome
 RT patients at cysteine residues in EGF-like domains.";

RL Hum. Mutat. 1:366-374(1992).
 RN [14]
 RP VARIANT MFS SER-2307.
 RX MEDLINE=92235290; PubMed=1569206;
 RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J. Jr.,
 RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;
 RT "Marfan phenotype variability in a family segregating a missense
 RT mutation in the epidermal growth factor-like motif of the fibrillin
 RL gene."; *J. Clin. Invest.* 89:1674-1680(1992).
 RN [15]
 RP VARIANTS MFS ILE-548 AND ALA-723.
 RX MEDLINE=94010946; PubMed=8406497;
 RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,
 RA Pyeritz R.E., Francomano C.A.;
 RA "Four novel FBN1 mutations: significance for mutant transcript level
 RT and EGF-like domain calcium binding in the pathogenesis of Marfan
 RT syndrome."; *Genomics* 17:468-475(1993).
 RL [16]
 RP VARIANTS MFS SER-2144.
 RX MEDLINE=93278402; PubMed=8504310;
 RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;
 RA "A novel fibrillin mutation in the Marfan syndrome which could
 RT disrupt calcium binding of the epidermal growth factor-like module."; *Hum. Mol. Genet.* 2:475-477(1993).
 RL [17]
 RP VARIANTS MFS R-862; Y-1117; P-1137 AND F-1589, AND VARIANT A-1148.
 RX MEDLINE=94108431; PubMed=8281141;
 RA Tynan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,
 RA Berg M.A., Miller D.C., Francke U.;
 RA "Mutation screening of complete fibrillin-1 coding sequence: report
 RT of five new mutations, including two in 8-cysteine domains."; *Hum. Mol. Genet.* 2:1813-1821(1993).
 RL [18]
 RP VARIANTS MFS GLY-217 AND ARG-2627.
 RX MEDLINE=95067970; PubMed=7977366;
 RA Karttunen L., Raghunath M., Loengqvist L., Peltonen L.;
 RA "A compound heterozygous Marfan patient: two defective fibrillin
 RT alleles result in a lethal phenotype."; *Am. J. Hum. Genet.* 55:1083-1091(1994).
 RL [19]
 RP VARIANT EL LYS-2447.
 RX MEDLINE=94245249; PubMed=8188302;
 RA Longqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,
 RA Peltonen L.;
 RA "A novel mutation of the fibrillin gene causing ectopia lentis."; *Genomics* 19:573-576(1994).
 RL [20]
 RP VARIANT MFS CYS-627.
 RX MEDLINE=94272487; PubMed=8004112;
 RA Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;
 RA "Two novel mutations and a neutral polymorphism in EGF-like domains
 RT of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan
 RT syndrome patients."; *Hum. Mol. Genet.* 3:373-375(1994).
 RL [21]
 RP VARIANT MFS CYS-122.
 RX MEDLINE=94314977; PubMed=8040326;
 RA Stahl-Hallengren C., Ukkonen T., Kainulainen K., Kristofersson U.,
 RA Saxne T., Tornqvist K., Peltonen L.;
 RA "An extra cysteine in one of the non-calcium-binding epidermal growth
 RT factor-like motifs of the FBN1 polypeptide is connected to a novel
 RT variant of Marfan syndrome."; *J. Clin. Invest.* 94:709-713(1994).
 RL [22]
 RP VARIANT MFS TYR-1223.
 RX MEDLINE=94351682; PubMed=8071963;
 RA Hewett D.R., Lynch J.R., Child A., Sykes B.C.;
 RA "A new missense mutation of fibrillin in a patient with Marfan
 RT syndrome."; *J. Med. Genet.* 31:338-339(1994).
 RL [23]

[illegible]

```

FT REPEAT 1957 1987 ANK 3.
FT REPEAT 1991 2020 ANK 4.
FT REPEAT 2024 2053 ANK 5.
FT REPEAT 2057 2086 ANK 6.
FT DISULFID 22 35 BY SIMILARITY.
FT DISULFID 29 45 BY SIMILARITY.
FT DISULFID 47 56 BY SIMILARITY.
FT DISULFID 62 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 128 BY SIMILARITY.
FT DISULFID 130 139 BY SIMILARITY.
FT DISULFID 145 156 BY SIMILARITY.
FT DISULFID 150 165 BY SIMILARITY.
FT DISULFID 167 176 BY SIMILARITY.
FT DISULFID 183 194 BY SIMILARITY.
FT DISULFID 188 203 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY.
FT DISULFID 226 242 BY SIMILARITY.
FT DISULFID 244 253 BY SIMILARITY.
FT DISULFID 260 271 BY SIMILARITY.
FT DISULFID 265 280 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT DISULFID 298 311 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 338 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 375 386 BY SIMILARITY.
FT DISULFID 380 397 BY SIMILARITY.
FT DISULFID 399 408 BY SIMILARITY.
FT DISULFID 415 428 BY SIMILARITY.
FT DISULFID 422 437 BY SIMILARITY.
FT DISULFID 439 448 BY SIMILARITY.
FT DISULFID 455 466 BY SIMILARITY.
FT DISULFID 460 475 BY SIMILARITY.
FT DISULFID 477 486 BY SIMILARITY.
FT DISULFID 493 504 BY SIMILARITY.
FT DISULFID 498 513 BY SIMILARITY.
FT DISULFID 515 524 BY SIMILARITY.
FT DISULFID 531 542 BY SIMILARITY.
FT DISULFID 536 551 BY SIMILARITY.
FT DISULFID 553 562 BY SIMILARITY.
FT DISULFID 559 579 BY SIMILARITY.
FT DISULFID 574 588 BY SIMILARITY.
FT DISULFID 590 599 BY SIMILARITY.
FT DISULFID 606 617 BY SIMILARITY.
FT DISULFID 611 626 BY SIMILARITY.
FT DISULFID 628 637 BY SIMILARITY.
FT DISULFID 644 654 BY SIMILARITY.
FT DISULFID 649 663 BY SIMILARITY.
FT DISULFID 665 674 BY SIMILARITY.
FT DISULFID 681 692 BY SIMILARITY.
FT DISULFID 703 712 BY SIMILARITY.
FT DISULFID 719 729 BY SIMILARITY.
FT DISULFID 724 738 BY SIMILARITY.
FT DISULFID 740 749 BY SIMILARITY.
FT DISULFID 756 767 BY SIMILARITY.
FT DISULFID 761 776 BY SIMILARITY.
FT DISULFID 778 787 BY SIMILARITY.
FT DISULFID 794 805 BY SIMILARITY.
FT DISULFID 799 814 BY SIMILARITY.
FT DISULFID 816 825 BY SIMILARITY.
FT DISULFID 832 843 BY SIMILARITY.
FT DISULFID 837 854 BY SIMILARITY.
FT DISULFID 856 865 BY SIMILARITY.
FT DISULFID 872 883 BY SIMILARITY.
FT DISULFID 877 892 BY SIMILARITY.
FT DISULFID 894 903 BY SIMILARITY.

```

```

FT DISULFID 910 921 BY SIMILARITY.
FT DISULFID 915 930 BY SIMILARITY.
FT DISULFID 932 941 BY SIMILARITY.
FT DISULFID 966 997 BY SIMILARITY.
FT DISULFID 991 1006 BY SIMILARITY.
FT DISULFID 1008 1017 BY SIMILARITY.
FT DISULFID 1024 1035 BY SIMILARITY.
FT DISULFID 1029 1044 BY SIMILARITY.
FT DISULFID 1046 1055 BY SIMILARITY.
FT DISULFID 1062 1073 BY SIMILARITY.
FT DISULFID 1067 1082 BY SIMILARITY.
FT DISULFID 1084 1093 BY SIMILARITY.
FT DISULFID 1100 1121 BY SIMILARITY.
FT DISULFID 1115 1130 BY SIMILARITY.
FT DISULFID 1132 1141 BY SIMILARITY.
FT DISULFID 1148 1159 BY SIMILARITY.
FT DISULFID 1153 1168 BY SIMILARITY.

Query Match 17.0%; Score 113; DB 1; Length 2524;
Best Local Similarity 34.3%; Pred. No. 0.01;
Matches 34; Conservative 5; Mismatches 42; Indels 18; Gaps 6;

QY 23 CRECY---CHCT---GGADCTGCTDCTGCGNCPNAHTCTDSKNCVNAATC-----TG 69
Db 959 CTFDVSNTCTCQPGRSGHICESNFPDCTE--SSCFNGGTCTDIDG---INTFTCQCPFTG 1014
QY 70 STKCNARTCTNSKDFEAKTCTDSTNCKATACINSTG 108
Db 1015 STCQHDINEC-DSKPLNGCTCDSYGYKCYCPQGYTG 1052

RESULT 9
KRUH_HUMAN
ID KRUH_HUMAN STANDARD; PRT; 169 AA.
AC P26371;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, ultra high-sulfur matrix protein A (UHS keratin A) (UHS
DE kera).
GN KRNI OR UHSK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=follicle;
RX MEDLINE=91115951; PubMed=1703541;
RA McKinnon P.J., Powell B.C., Rogers G.E.;
RT "Structure and expression of genes for a class of cysteine-rich
RT proteins of the cuticle layers of differentiating wool and hair
RT follicles.";
RT J. Cell Biol. 111:2587-2600(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148005; PubMed=10023043;
RA Perez C., Auriool J., Gerst C., Bernard B.A., Egly J.-M.;
RT "Genomic organization and promoter characterization of two human UHS
RT keratin genes.";
RL Gene 227:137-148(1999).
CC -! FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).
CC -! TISSUE SPECIFICITY: CUTICLE LAYERS OF DIFFERENTIATING
CC HAIR FOLLICLES.
CC -! DEVELOPMENTAL STAGE: AT A LATE STAGE OF FIBER DIFFERENTIATION.
CC -! DOMAIN: MAINLY COMPOSED OF CYS-RICH (CR), GLY-RICH (GR) AND SER-
CC RICH (SR) REPEATS.
CC -! SIMILARITY: BELONGS TO THE UHS KERATIN FAMILY.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X55293; CA39005.1; -.
DR EMBL: AJ006693; CA07189.1; -.
DR HSSP: P04355; 2MRT.
DR MIM: 148021; -.
KM Keratin; Repeat; Multigene family.
SQ SEQUENCE 169 AA; 16276 MW; 219B14FEEB49DAB CRC64;

Query Match 16.9%; Score 112.5; DB 1; Length 169;
Best Local Similarity 26.5%; Pred. No. 0.0016;
Matches 27; Conservative 14; Mismatches 44; Indels 17; Gaps 4;

QY 19 VMLCTECYCHCNGAGACTCTACTGCGCNCNPHNCTDCKNCVKATCTGS----- 70
DB 45 VCCVPACSCSGCKRCGCGSGGCGSC- GCGSCGCKPCCSSGCGSGCGSCGCK 103
QY 71 ---TKCNTARTCTNSKDCFEAKTCTDSTNCYKATCTACTSTGC 109
DB 104 PYCSCGCKRCCSSG--RGSSCGGSSCC---KPCSSSGC 140

RESULT 10
FBNL_BOVIN STANDARD; PRT; 2871 AA.
ID P88133;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibrillin 1 precursor (M340).
GN FBN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin.
RX MEDLINE=95137597; PubMed=7835900;
RA Talstra D.J., Potter K.A., Byers P.H.;
RT "Sequence of the coding region of the bovine fibrillin cDNA and
RT localization to bovine chromosome 10.";
RL Genomics 23:480-485(1994).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=96132851; PubMed=8557636;
RA Gibson M.A., Hatzinikolas G., Kumaratilake J.S., Sandberg L.B.,
RA Nicholl J.K., Sutherland G.R., Cleary E.G.;
RT "Further characterization of proteins associated with elastic fiber
RT microfibrils including the molecular cloning of MAGP-2 (M25).";
RL J. Biol. Chem. 271:1096-1103(1996).
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -1- PFM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS.
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L28748; AA74122.1; -.
DR HSSP: P35555; 1ABJ.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; Asx-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR00212; TB.
DR Pfam: PF00008; EGF; 46.
DR Pfam: PF00683; TB; 9.
DR PRINTS: PR00010; EGFBL00D.
DR SMART: SM00179; EGF_CA; 42.
DR SMART: SM00001; EGF_Like; 4.
DR PROSITE: PS00010; ASX_HYDROXYL; 43.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 38.
DR PROSITE: PS01187; EGF_CA; 45.
KM Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 1 27
FT CHAIN 28 2871
FT DOMAIN 81 112
FT DOMAIN 115 146
FT DOMAIN 147 178
FT DOMAIN 246 287
FT DOMAIN 288 329
FT REPEAT 330 390
FT REPEAT 392 446
FT DOMAIN 449 489
FT DOMAIN 490 529
FT DOMAIN 530 571
FT DOMAIN 572 612
FT DOMAIN 613 653
FT REPEAT 722 764
FT DOMAIN 765 806
FT DOMAIN 807 846
FT DOMAIN 910 951
FT REPEAT 1027 1078
FT DOMAIN 1079 1112
FT DOMAIN 1113 1154
FT DOMAIN 1155 1196
FT DOMAIN 1197 1237
FT DOMAIN 1238 1279
FT DOMAIN 1280 1321
FT DOMAIN 1322 1362
FT DOMAIN 1363 1403
FT DOMAIN 1404 1445
FT DOMAIN 1446 1486
FT DOMAIN 1487 1527
FT REPEAT 1528 1605
FT DOMAIN 1606 1647
FT DOMAIN 1648 1688
FT REPEAT 1689 1765
FT DOMAIN 1766 1807
FT DOMAIN 1808 1848
FT DOMAIN 1849 1890
FT DOMAIN 1891 1929
FT DOMAIN 1930 1972
FT DOMAIN 1973 2012
FT DOMAIN 2013 2054
FT REPEAT 2055 2126
FT DOMAIN 2127 2165
FT DOMAIN 2166 2205
FT DOMAIN 2206 2246
FT DOMAIN 2247 2290
FT DOMAIN 2291 2332
FT REPEAT 2333 2401
FT DOMAIN 2402 2443

POTENTIAL.
FIBRILLIN 1.
EGF-LIKE 1, NON-CALCIUM BINDING.
EGF-LIKE 2, NON-CALCIUM BINDING.
EGF-LIKE 3, NON-CALCIUM BINDING.
EGF-LIKE 4, CALCIUM-BINDING.
EGF-LIKE 5, CALCIUM-BINDING.
TGFBP 1.
PRO-RICH.
EGF-LIKE 6, NON-CALCIUM BINDING.
EGF-LIKE 7, CALCIUM-BINDING.
EGF-LIKE 8, CALCIUM-BINDING.
EGF-LIKE 9, CALCIUM-BINDING.
EGF-LIKE 10, CALCIUM-BINDING.
TGFBP 2.
EGF-LIKE 11, CALCIUM-BINDING.
EGF-LIKE 12, CALCIUM-BINDING.
EGF-LIKE 13, CALCIUM-BINDING.
EGF-LIKE 14, CALCIUM-BINDING.
TGFBP 3.
EGF-LIKE 15, CALCIUM-BINDING.
EGF-LIKE 16, CALCIUM-BINDING.
EGF-LIKE 17, CALCIUM-BINDING.
EGF-LIKE 18, CALCIUM-BINDING.
EGF-LIKE 19, CALCIUM-BINDING.
EGF-LIKE 20, CALCIUM-BINDING.
EGF-LIKE 21, CALCIUM-BINDING.
EGF-LIKE 22, CALCIUM-BINDING.
EGF-LIKE 23, CALCIUM-BINDING.
EGF-LIKE 24, CALCIUM-BINDING.
EGF-LIKE 25, CALCIUM-BINDING.
EGF-LIKE 26, CALCIUM-BINDING.
TGFBP 4.
EGF-LIKE 27, CALCIUM-BINDING.
TGFBP 5.
EGF-LIKE 28, CALCIUM-BINDING.
TGFBP 6.
EGF-LIKE 29, CALCIUM-BINDING.
EGF-LIKE 30, CALCIUM-BINDING.
EGF-LIKE 31, CALCIUM-BINDING.
EGF-LIKE 32, CALCIUM-BINDING.
EGF-LIKE 33, CALCIUM-BINDING.
EGF-LIKE 34, CALCIUM-BINDING.
EGF-LIKE 35, CALCIUM-BINDING.
TGFBP 7.
EGF-LIKE 36, CALCIUM-BINDING.
EGF-LIKE 37, CALCIUM-BINDING.
EGF-LIKE 38, CALCIUM-BINDING.
EGF-LIKE 39, CALCIUM-BINDING.
EGF-LIKE 40, CALCIUM-BINDING.
TGFBP 7.
EGF-LIKE 41, CALCIUM-BINDING.

FT DOMAIN 2444 2484 EGF-LIKE 42, CALCIUM-BINDING.
 FT DOMAIN 2485 2523 EGF-LIKE 43, CALCIUM-BINDING.
 FT DOMAIN 2524 2566 EGF-LIKE 44, CALCIUM-BINDING.
 FT DOMAIN 2567 2606 EGF-LIKE 45, CALCIUM-BINDING.
 FT DOMAIN 2607 2647 EGF-LIKE 46, CALCIUM-BINDING.
 FT DOMAIN 2648 2687 EGF-LIKE 47, CALCIUM-BINDING.
 FT DISULFID 85 94 BY SIMILARITY.
 FT DISULFID 89 100 BY SIMILARITY.
 FT DISULFID 102 111 BY SIMILARITY.
 FT DISULFID 119 129 BY SIMILARITY.
 FT DISULFID 123 134 BY SIMILARITY.
 FT DISULFID 136 145 BY SIMILARITY.
 FT DISULFID 150 160 BY SIMILARITY.
 FT DISULFID 154 166 BY SIMILARITY.
 FT DISULFID 168 177 BY SIMILARITY.
 FT DISULFID 250 262 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 FT DISULFID 273 286 BY SIMILARITY.
 FT DISULFID 292 304 BY SIMILARITY.
 FT DISULFID 299 313 BY SIMILARITY.
 FT DISULFID 315 328 BY SIMILARITY.
 FT DISULFID 453 465 BY SIMILARITY.
 FT DISULFID 460 474 BY SIMILARITY.
 FT DISULFID 476 488 BY SIMILARITY.
 FT DISULFID 494 504 BY SIMILARITY.
 FT DISULFID 499 513 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 534 546 BY SIMILARITY.
 FT DISULFID 541 555 BY SIMILARITY.
 FT DISULFID 557 570 BY SIMILARITY.
 FT DISULFID 576 587 BY SIMILARITY.
 FT DISULFID 582 596 BY SIMILARITY.
 FT DISULFID 598 611 BY SIMILARITY.
 FT DISULFID 617 628 BY SIMILARITY.
 FT DISULFID 623 637 BY SIMILARITY.
 FT DISULFID 639 652 BY SIMILARITY.
 FT DISULFID 727 739 BY SIMILARITY.
 FT DISULFID 734 748 BY SIMILARITY.
 FT DISULFID 750 763 BY SIMILARITY.
 FT DISULFID 769 781 BY SIMILARITY.
 FT DISULFID 776 790 BY SIMILARITY.
 FT DISULFID 792 805 BY SIMILARITY.
 FT DISULFID 811 821 BY SIMILARITY.
 FT DISULFID 816 830 BY SIMILARITY.
 FT DISULFID 832 845 BY SIMILARITY.
 FT DISULFID 914 926 BY SIMILARITY.
 FT DISULFID 921 935 BY SIMILARITY.
 FT DISULFID 937 950 BY SIMILARITY.
 FT DISULFID 1032 1044 BY SIMILARITY.
 FT DISULFID 1039 1053 BY SIMILARITY.
 FT DISULFID 1055 1068 BY SIMILARITY.
 FT DISULFID 1074 1086 BY SIMILARITY.
 FT DISULFID 1081 1095 BY SIMILARITY.
 FT DISULFID 1097 1111 BY SIMILARITY.
 FT DISULFID 1117 1129 BY SIMILARITY.
 FT DISULFID 1124 1138 BY SIMILARITY.
 FT DISULFID 1140 1153 BY SIMILARITY.
 FT DISULFID 1159 1171 BY SIMILARITY.
 FT DISULFID 1166 1180 BY SIMILARITY.
 FT DISULFID 1182 1195 BY SIMILARITY.
 FT DISULFID 1201 1212 BY SIMILARITY.
 FT DISULFID 1223 1236 BY SIMILARITY.
 FT DISULFID 1238 1254 BY SIMILARITY.
 FT DISULFID 1249 1263 BY SIMILARITY.
 FT DISULFID 1265 1278 BY SIMILARITY.
 FT DISULFID 1284 1296 BY SIMILARITY.
 FT DISULFID 1291 1305 BY SIMILARITY.
 FT DISULFID 1307 1320 BY SIMILARITY.
 FT DISULFID 1326 1339 BY SIMILARITY.
 FT DISULFID 1333 1348 BY SIMILARITY.
 FT DISULFID 1350 1361 BY SIMILARITY.
 FT DISULFID 1367 1380 BY SIMILARITY.

FT DISULFID 1374 1389 BY SIMILARITY.
 FT DISULFID 1391 1402 BY SIMILARITY.
 FT DISULFID 1408 1420 BY SIMILARITY.
 FT DISULFID 1415 1429 BY SIMILARITY.
 FT DISULFID 1431 1444 BY SIMILARITY.
 FT DISULFID 1450 1461 BY SIMILARITY.
 FT DISULFID 1456 1470 BY SIMILARITY.
 FT DISULFID 1472 1485 BY SIMILARITY.
 FT DISULFID 1491 1502 BY SIMILARITY.
 FT DISULFID 1497 1511 BY SIMILARITY.
 FT DISULFID 1513 1526 BY SIMILARITY.
 FT DISULFID 1610 1622 BY SIMILARITY.

Query Match 16.6%; Score 110.5; DB 1; Length 2871;
 Best Local Similarity 26.3%; Pred. No. 0.018;
 Matches 35; Conservative 10; Mismatches 41; Indels 47; Gaps 7;

QY 23 CT-----ECYCH-----CTGADCTCTDCTAG--CGNCPNAHTC----- 55
 DB 1212 CTNSECYSQSGPGLMPDQRCSTHDECDNPNICGGGCTNIPGNYRCLCYDGFMA 1271
 QY 56 -TDSKNCVAAATCTGSKTARTCTNSKDCF-----EAKT-CTD-----ST 95
 DB 1272 SEDMKTCVUNCDLNPNICLSTGTCENTKGSFICHODMGSGKKGTCTDINECIGAH 1331
 QY 96 NCYKATACNSTG 108
 DB 1332 NCDRAVCTNTAG 1344

RESULT 11
 FBNI_PIG ID FBNI_PIG STANDARD; PRT: 2871 AA.
 AC 09TV36;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Fibrillin 1 precursor.
 GN FBNI.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=99156858; PubMed=10036187;
 RA Biery N.J.; Eldadah Z.A.; Moore C.S.; Stetten G.; Spencer F.;
 RA Dietz H.C.;
 RT *Revised genomic organization of FBNI and significance for regulated
 gene expression*.
 RL Genomics 56:70-77(1999).
 CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
 THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
 LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
 CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
 FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
 MICROFIBRILS.
 CC -!- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF073800; AAD50328.1; -.
 DR HSP: P35555; IAPJ.
 DR InterPro: IPR000152; Asx_hydroxyl.

FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.
FT DISULFID 1674 1687 BY SIMILARITY.
FT DISULFID 1770 1782 BY SIMILARITY.
FT DISULFID 1777 1791 BY SIMILARITY.
FT DISULFID 1793 1806 BY SIMILARITY.
FT DISULFID 1812 1824 BY SIMILARITY.
FT DISULFID 1818 1833 BY SIMILARITY.
FT DISULFID 1835 1847 BY SIMILARITY.

Query Match 16.6%; Score 110.5; DB 1; Length 2871;

Best Local Similarity 26.3%; Pred. No. 0.018; 41; Indels 47; Gaps 7;

Matches 35; Conservative 10; Mismatches 41; Indels 47; Gaps 7;

QY 23 CT-----ECYCH-----CTGGADCTCTDCTACTG--CGCNPNAHTC-----55

DB 1212 CTNSESSEYSCQGFALMPDQSCDTIDCEDNPNICDGGCCTNIPGEYRCICYGFM 1271

QY 56 -TDSKNVKAATCTGSKTCNTARTCTNSKCF-----EART-CTD-----ST 95

DB 1272 SEDMKTCVDNECDLNPICLSCGTCENTKGSFICDMDYSGRKKTGCTDINECEIGAH 1331

QY 96 NCYKATCTNSTG 108

DB 1332 NCDRAVCTNTAG 1344

RESULT 12

MT_DREPO STANDARD; PRT; 73 AA.

AC 094550; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Metallothionein.

OS Dreissena polymorpha (Zebra mussel).

OC Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconcha; Veneroida;

OC Dreissenoida; Dreissenidae; Dreissena.

NCBI_TaxID=45954;

ON [1]

RP SEQUENCE FROM N.A.

RA Sczekan S.R., Engelken J., Hildebrandt A.;

RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 2-40.

RA Sczekan S.R., Engelken J., Hildebrandt A.;

RL Submitted (JUL-1997) to the SWISS-PROT data bank.

CC -1- FUNCTION: THE METALLOTHIONEINS ARE INVOLVED IN THE CELLULAR

CC SEQUESTRATION OF TOXIC METAL IONS (BY SIMILARITY).

CC -1- INDUCTION: BY CADMIUM.

CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U67347; AAB07548.1; -.

DR HSSP; P02795; ZMHU.

DR InterPro; IPRO01008; Metallthion_mollusc.

DR PRINTS; PR00875; ITMOLLUSC.

KW Metal-binding; Metal-thiolate cluster; Chelation; Cadmium.

SO SEQUENCE 73 AA; 7364 MW; DD3398705C4DAEE9 CRC64;

Query Match 16.4%; Score 109; DB 1; Length 73;

Best Local Similarity 28.6%; Pred. No. 0.0017;

Matches 20; Conservative 6; Mismatches 40; Indels 4; Gaps 1;

QY 28 CHTGGADCTCTDCTGCGNPNNAHTCTDSCNKVKAATCTGSKTCNTARTCTNSKCFE 87

DB 5 CMCVETGDCRCADGSCSDSCSNC---KCGDSCCKSPNCCGKNVTKCGEMCQCGVGTG 60

QY 88 AKTCTDSTNC 97

DB 61 PDSCTCDSC 70

RESULT 13

ITB7_HUMAN STANDARD; PRT; 798 AA.

AC P26010; 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Integrin beta-7 precursor.

GN ITGB7.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_TaxID=9606;

ON [1]

RP SEQUENCE FROM N.A. (SHORT AND LONG FORMS).

RC TISSUE=leukocyte;

RA MEDLINE=91250405; PubMed=2040616;

RA Erle D.J., Ruegg C., Sheppard D., Pytela R.;

RT "Complete amino acid sequence of an integrin beta subunit (beta 7)

RT identified in leukocytes.";

RL J. Biol. Chem. 266:11009-11016(1991).

RN [2]

RP SEQUENCE FROM N.A.

RA MEDLINE=91190778; PubMed=2083230;

RA Yuan Q., Jiang W.-M., Krissansen G.W., Watson J.D.;

RT "Cloning and sequence analysis of a novel beta 2-related integrin

RT transcript from T lymphocytes: homology of integrin cysteine-rich

RT repeats to domain III of laminin B chains.";

RL Int. Immunol. 2:1097-1108(1990).

RN [3]

RP REVISIONS.

RA MEDLINE=92135083; PubMed=1777426;

RA Yuan Q., Jiang W.-M., Krissansen G.W., Watson J.D.;

RT "Cloning and sequence analysis of a novel beta 2-related integrin

RT transcript from T lymphocytes: homology of integrin cysteine-rich

RT repeats to domain III of laminin B chains.";

RL Int. Immunol. 3:1373-1374(1991).

RN [4]

RP SEQUENCE FROM N.A.

RA MEDLINE=93002753; PubMed=1382574;

RA Jiang W.-M., Jenkins D., Yuan Q., Leung E., Choo K.H., Watson J.D.,

RA Krissansen G.W.;

RT "The gene organization of the human beta 7 subunit, the common beta

RT subunit of the leukocyte integrins HML-1 and IPAM-1.";

RL Int. Immunol. 4:1031-1040(1992).

RN [5]

RP MUTAGENESIS OF ASP-159.

RA MEDLINE=20400502; PubMed=10837471;

RA Higgins J.M.G., Cernadas M., Tan K., Irie A., Wang J.-H., Takada Y.,

RA Brenner M.B.;

RT "The role of alpha and beta chains in ligand recognition by beta 7

RT integrins.";

RL J. Biol. Chem. 275:25652-25664(2000).

CC -1- FUNCTION: INTEGRIN ALPHA-4/BETA-7 (PEYER'S PATCHES-SPECIFIC HOMING

CC RECEPTOR LPAM-1) IS EXPECTED TO PLAY A ROLE IN ADHESIVE

CC INTERACTIONS OF LEUCOCYTES. IT IS A RECEPTOR FOR FIBRONECTIN AND

CC RECOGNIZES ONE OR MORE DOMAINS WITHIN THE ALTERNATIVELY SPLICED

CC CS-1 REGION OF FIBRONECTIN. INTEGRIN ALPHA-4/BETA-7 IS ALSO A

CC RECEPTOR FOR MADCAM1 AND VCAM1. IT RECOGNIZES THE SEQUENCE L-D-T

CC IN MADCAM1. INTEGRIN ALPHA-E/BETA-7 (HML-1) IS A RECEPTOR FOR E-

CC CADHERIN.

CC -1- SUBUNIT: HETERO-DIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-7

CC ASSOCIATES WITH EITHER ALPHA-4 OR ALPHA-E.
 CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF LEUKOCYTE LINES.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 WFVA-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S80335; AAB23332.1; -;
 DR EMBL: M62880; AAA59185.1; -;
 DR EMBL: M68892; AAA59184.1; -;
 DR EMBL: S49378; AAB23688.1; -;
 DR EMBL: S49364; AAB23688.1; JOINED.
 DR EMBL: S49365; AAB23688.1; JOINED.
 DR EMBL: S49366; AAB23688.1; JOINED.
 DR EMBL: S49367; AAB23688.1; JOINED.
 DR EMBL: S49368; AAB23688.1; JOINED.
 DR EMBL: S49369; AAB23688.1; JOINED.
 DR EMBL: S49370; AAB23688.1; JOINED.
 DR EMBL: S49371; AAB23688.1; JOINED.
 DR EMBL: S49373; AAB23688.1; JOINED.
 DR EMBL: S49374; AAB23688.1; JOINED.
 DR EMBL: S49375; AAB23688.1; JOINED.
 DR EMBL: S49377; AAB23688.1; JOINED.
 DR EMBL: L23823; AAA36118.1; -;
 DR EMBL: L23810; AAA36118.1; JOINED.
 DR EMBL: L23811; AAA36118.1; JOINED.
 DR EMBL: L23812; AAA36118.1; JOINED.
 DR EMBL: L23813; AAA36118.1; JOINED.
 DR EMBL: L23814; AAA36118.1; JOINED.
 DR EMBL: L23815; AAA36118.1; JOINED.
 DR EMBL: L23816; AAA36118.1; JOINED.
 DR EMBL: L23817; AAA36118.1; JOINED.
 DR EMBL: L23818; AAA36118.1; JOINED.
 DR EMBL: L23819; AAA36118.1; JOINED.
 DR EMBL: L23820; AAA36118.1; JOINED.
 DR EMBL: L23822; AAA36118.1; JOINED.
 DR PIR: A40526; AA0526.
 DR MIM: 147559; -;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002369; Integrin_B.
 DR InterPro: IPR001169; Integrin_beta_C.
 DR InterPro: IPR003659; PSI.
 DR InterPro: IPR002035; WFVA.
 DR Pfam: PF00362; Integrin_B.1.
 DR PRINTS: PR01186; INTEGRINB.
 DR PRODOM: PD001811; Integrin_B.1.
 DR SMART: SM00001; EGF_Like; 1.
 DR SMART: SM00187; INB; 1.
 DR SMART: SM00423; PST; 1.
 DR SMART: SM00327; VWA; 1.
 DR PROSITE: PS00243; INTEGRIN_BETA_3.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_4.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 KM Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
 KM Repeat; Signal; Phosphorylation; Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 798 INTEGRIN BETA-7
 FT DOMAIN 20 723 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 724 746 POTENTIAL.
 FT DOMAIN 747 798 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 150 389 WFVA-LIKE.
 FT DOMAIN 478 640 4 CYSTEINE-RICH TANDEM REPEATS.
 FT REPEAT 478 526 I.
 FT REPEAT 527 565 II.

FT REPEAT 566 604 III.
 FT REPEAT 605 640 IV.
 FT DISULFID 45 476 BY SIMILARITY.
 FT DISULFID 51 61 BY SIMILARITY.
 FT DISULFID 54 91 BY SIMILARITY.
 FT DISULFID 64 80 BY SIMILARITY.
 FT DISULFID 216 223 BY SIMILARITY.
 FT DISULFID 271 311 BY SIMILARITY.
 FT DISULFID 412 428 BY SIMILARITY.
 FT DISULFID 448 478 BY SIMILARITY.
 FT DISULFID 474 478 BY SIMILARITY.
 FT DISULFID 488 500 BY SIMILARITY.
 FT DISULFID 497 537 BY SIMILARITY.
 FT DISULFID 502 511 BY SIMILARITY.
 FT DISULFID 513 527 BY SIMILARITY.
 FT DISULFID 543 548 BY SIMILARITY.
 FT DISULFID 545 574 BY SIMILARITY.
 FT DISULFID 550 559 BY SIMILARITY.
 FT DISULFID 561 566 BY SIMILARITY.
 FT DISULFID 580 585 BY SIMILARITY.
 FT DISULFID 582 613 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 598 605 BY SIMILARITY.
 FT DISULFID 619 624 BY SIMILARITY.
 FT DISULFID 621 666 BY SIMILARITY.
 FT DISULFID 626 635 BY SIMILARITY.
 FT DISULFID 638 641 BY SIMILARITY.
 FT DISULFID 645 654 BY SIMILARITY.
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 674 674 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD_RES 778 778 PHOSPHORYLATION (BY TYR-KINASES)
 FT (BY SIMILARITY).
 FT VARSPLIC 501 648 MISSING (IN SHORT ISOFORM).
 FT MUTAGEN 159 159 D->A: LOSS OF INTEGRIN ALPHA-E/BETA-7
 FT BINDING TO E-CADHERIN AND OF INTEGRIN
 FT ALPHA-4/BETA-7 BINDING TO MADCAM1.
 SQ SEQUENCE 798 AA; 86903 MW; CBE275E0E9992385 CRC64;
 Query Match 16.4%; Score 109; DB 1; Length 798;
 Best Local Similarity 32.0%; Pred. No. 0.0098;
 Matches 32; Conservative 5; Mismatches 47; Indels 16; Gaps 5;
 QY 21 CLCTECYCH-----CTGGADCTSCD---ACTGCGNCP-NAHTCTDSKNVKAATCT 68
 DB 580 CCGCVCHCHNARTGRACSCSGDMSICSPGGICSGHGRCKNRCQCLDYGALCDQCP 639
 QY 69 G-STKCNARTCTNSKDCFEAKTCTDSTNCKYKATCTST 107
 DB 640 GCKTPCRHRDCA---ECGAFRTGLPLATNCSTACHTVNT 676
 RESULT 14
 C170_GIATA STANDARD; PRT; 328 AA.
 AC P15799;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Surface antigen CRP170 (Fragment).
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.
 OX NCBI_TaxID=5741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 30957 / WB;
 RX MEDLINE=88089405; PubMed=3335828;
 RA Adam R.D., Aggarwal A., Lal A.A., de la Cruz V.F., McCutchan T.,

RA Nash T.E.;
RT "Antigenic variation of a cysteine-rich protein in Giardia lamblia."; J. Exp. Med. 167:109-118(1988).
CC -1- MISCELLANEOUS: CYSTEINE-RICH, ANTIGENICALLY VARIANT SURFACE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X06741; CAA29916.1; -.
DR PIR: S00530; S00530.
DR InterPro: IPR002174; Furin-like.
DR SMART: SM00261; FU; 3.
DR Repeat: Antigen.
FT NON_TER 1 1
FT REPEAT 38 102
FT REPEAT 103 167
FT NON_TER 328 328
SO SEQUENCE 328 AA; 33438 MW; 373A697A30EDCA21 CRC64;

Query Match 16.3%; Score 108.5; DB 1; Length 328;
Best Local Similarity 28.5%; Pred. No. 0.0057;
Matches 37; Conservative 9; Mismatches 41; Indels 43; Gaps 7;

Qy 23 CTECYCHCTGAGADCTCTDA-----CTGC-----GNCPPNA---HT---CT 56
Db 167 CAEGCKPMTAGTQGFSCDANCERCDDNDVCARCSGAPENGACPAEGCHSCGCGCT 226
Qy 57 DSKNCVRAKATCTG-----STGCNTRARCTCNKDCFEAKCTCDSTN-CYKA 100
Db 227 ENAMTNQADKCTGCKEGRYLKPESAAQSGALTAEBCTSDKTHFTREKAGDSKMGCLSC 286
Qy 101 TACTNS-TGC 109
Db : 1: 111
Db 287 SDATHGTTGC 296

RESULT 15
FBN1_MOUSE STANDARD; PRT; 2871 AA.
ID FBN1_MOUSE
AC 061554; 060826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Fibrillin 1 precursor.
GN FBN1 OR FBN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130561; PubMed=7829516;
RA Yin W., Germiller J., Sanguinetti C., Smiley E., Pangllian T.,
RA Pereira L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene."; J. Biol. Chem. 270:1798-1806(1995).
RL J. Biol. Chem. 270:1798-1806(1995).
RM [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RA Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -1- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE

CC MICROFIBRILS (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U29454; AAA56840.1; -.
DR EMBL: U22493; AAA64217.1; -.
DR HSSP: P35555; IAPJ.
DR WGI: 95489; FBN1.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002212; TB.
DR Pfam: PF00008; EGF; 46.
DR Pfam: PF00683; TB; 9.
DR SMART: SM00179; EGF_CA; 42.
DR SMART: SM00001; EGF_Like; 4.
DR PROSITE: PS00010; ASX_HYDROXYL; 43.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 38.
DR PROSITE: PS01187; EGF_CA; 45.
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 1 27
FT CHAIN 28 2871
FT DOMAIN 81 112
FT DOMAIN 115 146
FT DOMAIN 147 178
FT DOMAIN 246 287
FT DOMAIN 288 329
FT DOMAIN 330 401
FT DOMAIN 402 446
FT DOMAIN 449 489
FT DOMAIN 490 529
FT DOMAIN 530 571
FT DOMAIN 572 612
FT DOMAIN 613 653
FT DOMAIN 656 721
FT DOMAIN 723 764
FT DOMAIN 765 806
FT DOMAIN 807 846
FT DOMAIN 910 951
FT DOMAIN 952 1018
FT DOMAIN 1028 1069
FT DOMAIN 1070 1112
FT DOMAIN 1113 1154
FT DOMAIN 1155 1196
FT DOMAIN 1197 1237
FT DOMAIN 1238 1279
FT DOMAIN 1280 1321
FT DOMAIN 1322 1362
FT DOMAIN 1363 1403
FT DOMAIN 1404 1445
FT DOMAIN 1446 1486
FT DOMAIN 1487 1527
FT DOMAIN 1528 1599
FT DOMAIN 1599 1647
FT DOMAIN 1648 1688
FT DOMAIN 1689 1758
FT DOMAIN 1766 1807
FT DOMAIN 1808 1848
FT DOMAIN 1849 1890
FT DOMAIN 1891 1929
FT DOMAIN 1930 1972
FT DOMAIN 1973 2012

FIBRILLIN 1.
EGF-LIKE 1, NON-CALCIUM BINDING.
EGF-LIKE 2, NON-CALCIUM BINDING.
EGF-LIKE 3, NON-CALCIUM BINDING.
EGF-LIKE 4, CALCIUM-BINDING.
EGF-LIKE 5, CALCIUM-BINDING.
TGFBP 1.
PRO-RICH.
EGF-LIKE 6, NON-CALCIUM BINDING.
EGF-LIKE 7, CALCIUM-BINDING.
EGF-LIKE 8, CALCIUM-BINDING.
EGF-LIKE 9, CALCIUM-BINDING.
EGF-LIKE 10, CALCIUM-BINDING.
TGFBP 2.
EGF-LIKE 11, CALCIUM-BINDING.
EGF-LIKE 12, CALCIUM-BINDING.
EGF-LIKE 13, CALCIUM-BINDING.
EGF-LIKE 14, CALCIUM-BINDING.
TGFBP 3.
EGF-LIKE 15, CALCIUM-BINDING.
EGF-LIKE 16, CALCIUM-BINDING.
EGF-LIKE 17, CALCIUM-BINDING.
EGF-LIKE 18, CALCIUM-BINDING.
EGF-LIKE 19, CALCIUM-BINDING.
EGF-LIKE 20, CALCIUM-BINDING.
EGF-LIKE 21, CALCIUM-BINDING.
EGF-LIKE 22, CALCIUM-BINDING.
EGF-LIKE 23, CALCIUM-BINDING.
EGF-LIKE 24, CALCIUM-BINDING.
EGF-LIKE 25, CALCIUM-BINDING.
EGF-LIKE 26, CALCIUM-BINDING.
TGFBP 4.
EGF-LIKE 27, CALCIUM-BINDING.
EGF-LIKE 28, CALCIUM-BINDING.
TGFBP 5.
EGF-LIKE 29, CALCIUM-BINDING.
EGF-LIKE 30, CALCIUM-BINDING.
EGF-LIKE 31, CALCIUM-BINDING.
EGF-LIKE 32, CALCIUM-BINDING.
EGF-LIKE 33, CALCIUM-BINDING.
EGF-LIKE 34, CALCIUM-BINDING.

```
FT DOMAIN 2013 2054 EGF-LIKE 35, CALCIUM-BINDING.
FT DOMAIN 2055 2121 TGFBRP 6.
FT DOMAIN 2127 2165 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2166 2205 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2206 2246 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2247 2290 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2291 2332 EGF-LIKE 40, CALCIUM-BINDING.
FT DOMAIN 2333 2400 TGFBRP 7.
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2485 2523 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2567 2606 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2648 2687 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 89 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
FT DISULFID 119 129 BY SIMILARITY.
FT DISULFID 123 134 BY SIMILARITY.
FT DISULFID 136 145 BY SIMILARITY.
FT DISULFID 150 160 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 250 262 BY SIMILARITY.
FT DISULFID 257 271 BY SIMILARITY.
FT DISULFID 273 286 BY SIMILARITY.
FT DISULFID 292 304 BY SIMILARITY.
FT DISULFID 299 313 BY SIMILARITY.
FT DISULFID 315 328 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 460 474 BY SIMILARITY.
FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 776 790 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
```

```
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.
```

Query Match 16.3% Score 108.5; DB 1; Length 2871;
Best Local Similarity 26.3%; Pred. No. 0.027;
Matches 35; Conservative 9; Mismatches 42; Indels 47; Gaps 7;

```
QY 23 CT-----ECYCH-----CTGADCTSCDTACTG--CGNCPNAHTC----- 55
DB 1212 CTNSDGSYECSQPGFALMPDQRSCDTIDQCEDNPNICDGGCTNINPEYRCICLYDFGMA 1271
QY 56 -TDSKNCKYKATCTGCTSCNTARTCTNSKDCF-----EAKT-CTP-----ST 95
DB 1272 SEDMKTQVYNECDLNPNICLSGTCTGCTGKSFICHCMDGYSGKKGTGCTDINECEIGAH 1331
QY 96 NCKYKATCTNSNG 108
DB 1332 NCGRHAVCNTNAG 1344
```

Search completed: October 21, 2002, 16:38:48
Job time : 16.1765 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:44 : Search time 36.2353 Seconds
(without alignments)
534.712 Million cell updates/sec

Title: US-10-032-658-11

Perfect score: 664

Sequence: 1 MAFKTCGFSEKMWLAVIVM.....DSTNCYKATACINSTGCPGH 112

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SPTEMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 664 | 100.0 | 112 | 5 | 016120 tenebrio mo |
| 2 | 659 | 99.2 | 112 | 5 | 09u746 tenebrio mo |
| 3 | 612 | 92.2 | 112 | 5 | 09u747 tenebrio mo |
| 4 | 603 | 90.8 | 124 | 5 | 016121 tenebrio mo |
| 5 | 599 | 90.2 | 112 | 5 | 016119 tenebrio mo |
| 6 | 599 | 90.2 | 124 | 5 | 09u744 tenebrio mo |
| 7 | 592 | 89.2 | 112 | 5 | 09u748 tenebrio mo |
| 8 | 583 | 87.8 | 112 | 5 | 09u745 tenebrio mo |
| 9 | 487 | 73.3 | 148 | 5 | 016122 tenebrio mo |
| 10 | 385 | 58.0 | 118 | 5 | 09nc97 dendroides |
| 11 | 385 | 58.0 | 122 | 5 | 09ncr4 dendroides |
| 12 | 377.5 | 56.9 | 108 | 5 | 046351 dendroides |
| 13 | 377.5 | 56.9 | 108 | 5 | 046352 dendroides |
| 14 | 372 | 56.0 | 136 | 5 | 09ncr2 dendroides |
| 15 | 364.5 | 54.9 | 123 | 5 | 09nc99 dendroides |
| 16 | 361 | 54.4 | 109 | 5 | 046346 dendroides |

| | | | | | |
|----|-------|------|------|----|--------------------|
| 17 | 359 | 54.1 | 121 | 5 | 09ncr1 dendroides |
| 18 | 357.5 | 53.8 | 129 | 5 | 09ncr0 dendroides |
| 19 | 357.5 | 53.8 | 148 | 5 | 09ncr8 dendroides |
| 20 | 329.5 | 49.6 | 98 | 5 | 09ncr6 dendroides |
| 21 | 327.5 | 49.3 | 91 | 5 | 09ncr3 dendroides |
| 22 | 326 | 49.1 | 104 | 5 | 09ncr5 dendroides |
| 23 | 325.5 | 49.0 | 96 | 5 | 096419 dendroides |
| 24 | 138.5 | 20.9 | 313 | 5 | 024330 dictyostel |
| 25 | 133 | 20.0 | 107 | 5 | 09ncg9 crassostrea |
| 26 | 124.5 | 18.8 | 548 | 5 | 096045 giardia lam |
| 27 | 120.5 | 18.1 | 350 | 5 | 094389 leishmania |
| 28 | 119.5 | 18.0 | 861 | 11 | 09QW58 mus sp. mot |
| 29 | 118.5 | 17.8 | 2189 | 5 | 09BI05 elmeria ten |
| 30 | 117.5 | 17.7 | 401 | 5 | 09NH87 giardia lam |
| 31 | 117 | 17.6 | 862 | 11 | 099JC2 mus musculu |
| 32 | 116.5 | 17.5 | 75 | 5 | 09U1N5 crassostrea |
| 33 | 116 | 17.5 | 667 | 5 | 09SW01 giardia lam |
| 34 | 113.5 | 17.1 | 548 | 5 | 021629 caenorhabdi |
| 35 | 113.5 | 17.1 | 1174 | 11 | 099K58 mus musculu |
| 36 | 112.5 | 16.9 | 100 | 5 | 096260 littorina l |
| 37 | 112.5 | 16.9 | 169 | 4 | 014564 homo sapien |
| 38 | 112.5 | 16.9 | 1210 | 6 | 095ND4 felis silve |
| 39 | 112 | 16.9 | 718 | 5 | 09BI07 entamoeba h |
| 40 | 112 | 16.9 | 1111 | 5 | 09XWD6 caenorhabdi |
| 41 | 111.5 | 16.8 | 1698 | 5 | 094438 chironomus |
| 42 | 111 | 16.7 | 966 | 5 | 022378 caenorhabdi |
| 43 | 111 | 16.7 | 1202 | 11 | P97607 rattus norv |
| 44 | 110.5 | 16.6 | 969 | 4 | 096K66 homo sapien |
| 45 | 110.5 | 16.6 | 2872 | 11 | 09WU08 rattus norv |

ALIGNMENTS

| RESULT 1 | ID | PRELIMINARY; | PRT; | 112 AA. |
|----------|---|--------------|--------------------|---|
| 016120 | 016120: | | | |
| DT | 01-JAN-1998 (TREMBLrel. 05, Created) | | | |
| DT | 01-JAN-1998 (TREMBLrel. 05, Last sequence update) | | | |
| DT | 01-DEC-2001 (TREMBLrel. 19, Last annotation update) | | | |
| DE | THERMAL HYSTERESIS PROTEIN ISOFORM YL-2 (1-3) PRECURSOR. | | | |
| OS | Tenebrio molitor (Yellow mealworm). | | | |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | | |
| OC | Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; | | | |
| OC | Cucujiformia; Tenebrionidae; Tenebrio. | | | |
| OX | NCBI_Taxid=7067; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | TISSUE=FATBODY; | | | |
| RX | MEDLINE=94400482; PubMed=10471292; | | | |
| RA | Liou Y.C., Thibault P., Walker V.K., Davies P.L.; | | | |
| RT | "A complex family of highly heterogeneous and internally repetitive | | | |
| RT | hyperactive antifreeze proteins from the beetle Tenebrio molitor."; | | | |
| RL | Biochemistry 38:11415-11424(1999). | | | |
| DR | EMBL; AF160495; AAB70751.1; | | | |
| DR | InterPro; IPR003460; AFP. | | | |
| DR | Pfam; PF02420; AFP; 8. | | | |
| KW | Signal. | | | |
| FT | SIGNAL | 1 | 28 | POTENTIAL. |
| FT | CHAIN | 29 | 112 | THERMAL HYSTERESIS PROTEIN ISOFORM YL-2 |
| FT | SEQUENCE | 112 AA; | 11667 MW; | 5883E6DCDF0F805 CRC64; |
| SQ | Query Match | 100.0%; | Score 664; | DB 5; |
| | Best Local Similarity | 100.0%; | Pred. No. 4.3e-67; | Length 112; |

August 21, 1997

| Query Match | Best Local Similarity | Matches 103; Conservative | Score 612; DB 5; Length 112; Pred. No. 2, 8e-61; Mismatches 5; Indels 0; Gaps 0; |
|-------------|--|---------------------------|--|
| 1 | 90.8%; | 92.28; | Score 612; DB 5; Length 112; Pred. No. 2, 8e-61; Mismatches 5; Indels 0; Gaps 0; |
| 1 | MAFPTCGSKRWLIIVAVIMVCLTECYCHCTGADCTSCDTACTGCGCNPAAHTCTDSK - 59 | 1 | MAFPTCGSKRWLIIVAVIMVCLTECYCHCTGADCTSCDTACTGCGCNPAAHTCTDSK 60 |
| 1 | MAFPTCGSKRWLIIVAVIMVCLTECYCHCTGADCTSCDTACTGCGCNPAAHTCTDSK 60 | 1 | MAFPTCGSKRWLIIVAVIMVCLTECYCHCTGADCTSCDTACTGCGCNPAAHTCTDSK 60 |
| 60 | -----NCVKATCTGCTKCNARFCTNSKDCFEAKTCTDSTNCKAKACTNSTG 108 | 61 | CVKATCTGCTKCNARFCTNSKDCFEAKTCTDSTNCKAKACTNSTG 112 |
| 61 | CVKATCTGCTKCNARFCTNSKDCFEAKTCTDSTNCKAKACTNSTG 108 | 61 | CVKATCTGCTKCNARFCTNSKDCFEAKTCTDSTNCKAKACTNSTG 112 |
| 109 | CPGH 112 | 109 | CPGH 112 |

Db 121 CPGH 124

RESULT 5

016119 PRELIMINARY; PRT; 112 AA.

AC 016119; 1

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-1 (2-14) PRECURSOR.

OS Tenebrio molitor (Yellow mealworm).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;

OC Cucujiformia; Tenebrionidae; Tenebrio.

OX NCBI_TaxID=7067;

RN 11

RP SEQUENCE FROM N.A.

RC TISSUE=FATBODY;

RX MEDLINE=97429943; PubMed=9285581;

RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;

RT "A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle Tenebrio molitor.";

RL Biochemistry 38:11415-11424(1999).

DR EMBL; AF160494; AAB7050.1; -.

DR InterPro; IPR003460; AFP.

DR Pfam; PF02420; AFP; 8.

KW Signal.

FT CHAIN 1 28 POTENTIAL.

FT CHAIN 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM YL-1

FT CHAIN (2-14).

SO SEQUENCE 112 AA; 11558 MW; C06D7AB6401BC909 CRC64;

Query Match 90.2%; Score 599; DB 5; Length 112;

Best Local Similarity 90.2%; Pred. No. 8.1e-60;

Matches 101; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAFKTCGFSKKMLVIAVIMCLCTECYCHCTGADCTSDTACTGCGNCPNAHTCTDSKN 60

DB 1 MAFKTCGFSKKMLVIAVIMCLCTECYCHCTGADCTSDTACTGCGNCPNAHTCTDSKN 60

QY 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

RESULT 6

090744 PRELIMINARY; PRT; 124 AA.

AC 090744; 1

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE THERMAL HYSTERESIS PROTEIN ISOFORM 5-15 PRECURSOR.

OS Tenebrio molitor (Yellow mealworm).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;

OC Cucujiformia; Tenebrionidae; Tenebrio.

OX NCBI_TaxID=7067;

RN 11

RP SEQUENCE FROM N.A.

RC TISSUE=FATBODY;

RX MEDLINE=99400482; PubMed=10471292;

RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;

RT "A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle Tenebrio molitor.";

RL Biochemistry 38:11415-11424(1999).

DR EMBL; AF159114; AAD55256.1; -.

DR InterPro; IPR003460; AFP.

DR Pfam; PF02420; AFP; 8.

KW Signal.

FT CHAIN 1 28 POTENTIAL.

FT CHAIN 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 4-9,

FT CHAIN SEQUENCE 112 AA; 11532 MW; 8A6124CD31DDE19 CRC64;

QY 1 MAFKTCGFSKKMLVIAVIMCLCTECYCHCTGADCTSDTACTGCGNCPNAHTCTDSKN 60

DB 1 MAFKTCGFSKKMLVIAVIMCLCTECYCHCTGADCTSDTACTGCGNCPNAHTCTDSKN 60

QY 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

RL Biochemistry 38:11415-11424(1999).

DR EMBL; AF159114; AAD55260.1; -.

DR InterPro; IPR003460; AFP.

DR Pfam; PF02420; AFP; 9.

KW Signal.

FT CHAIN 1 28 POTENTIAL.

FT CHAIN 29 124 THERMAL HYSTERESIS PROTEIN ISOFORM 5-15,

SO SEQUENCE 124 AA; 13005 MW; 8EC6871047F2890B CRC64;

Query Match 90.2%; Score 599; DB 5; Length 124;

Best Local Similarity 83.1%; Pred. No. 8.9e-60;

Matches 103; Conservative 3; Mismatches 6; Indels 12; Gaps 1;

QY 1 MAFKTCGFSKKMLVIAVIMCLCTECYCHCTGADCTSDTACTGCGNCPNAHTCTDSKN 53

DB 1 MAFKTCGFSKKMLVIAVIMCLCTECYCHCTGADCTSDTACTGCGNCPNAHTCTDSKN 53

QY 54 -----TCTDSKNCVKATCTGSGTKCNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTG 108

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 120

QY 109 CPGH 112

DB 121 CPGH 124

RESULT 7

090748 PRELIMINARY; PRT; 112 AA.

AC 090748; 1

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE THERMAL HYSTERESIS PROTEIN ISOFORM 4-9 PRECURSOR.

OS Tenebrio molitor (Yellow mealworm).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;

OC Cucujiformia; Tenebrionidae; Tenebrio.

OX NCBI_TaxID=7067;

RN 11

RP SEQUENCE FROM N.A.

RC TISSUE=FATBODY;

RX MEDLINE=99400482; PubMed=10471292;

RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;

RT "A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle Tenebrio molitor.";

RL Biochemistry 38:11415-11424(1999).

DR EMBL; AF159114; AAD55256.1; -.

DR InterPro; IPR003460; AFP.

DR Pfam; PF02420; AFP; 8.

KW Signal.

FT CHAIN 1 28 POTENTIAL.

FT CHAIN 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 4-9,

FT CHAIN SEQUENCE 112 AA; 11532 MW; 8A6124CD31DDE19 CRC64;

QY 1 MAFKTCGFSKKMLVIAVIMCLCTECYCHCTGADCTSDTACTGCGNCPNAHTCTDSKN 60

DB 1 MAFKTCGFSKKMLVIAVIMCLCTECYCHCTGADCTSDTACTGCGNCPNAHTCTDSKN 60

QY 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKAACTGSGKNTARTCTNSKCCFEAKTCTDSTNCTKATCTNSTGCPGH 112

RESULT 8

090745 PRELIMINARY; PRT; 112 AA.

AC 090745; 1

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM C-9 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF159117; AAD55259.1; -.
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 11.
KW Signal.
FT SIGNAL. 1 28 POTENTIAL.
FT CHAIN 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM C-9.
SQ SEQUENCE 112 AA; 12002 MW; A4CA9DFB0AF81FB CRC64;
Query Match 87.8%; Score 583; DB 5; Length 112;
Best Local Similarity 87.5%; Pred. No. 5e-58;
Matches 98; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
QY 1 MARKTCGFSKKMLVIAIVMCLCTECYCHCTGAGADCTGCTGAGCNCNPAHTCTDSKN 60
DB 1 MARKTCGFSKKMLVIAIVMCLCTECYCHCTGAGADCTGCTGAGCNCNPAHTCTDSKN 60
QY 61 CVRAATCTGTCNATCTCNSKDCFEAKTCTSTNCYKATCTACTNSTGCGPH 112
DB 61 CVRAATCTGTCNATCTCNSKDCFEAKTCTSTNCYKATCTACTNSTGCGPH 112
RESULT 9
ID 016122 PRELIMINARY; PRT; 148 AA.
AC 016122;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-4 (2-20) PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF160497; AAB70753.1; -.
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 11.
KW Signal.
FT SIGNAL. 1 28 POTENTIAL.
FT CHAIN 29 148 THERMAL HYSTERESIS PROTEIN ISOFORM YL-4
SQ SEQUENCE 148 AA; 15311 MW; EF07B36BB8684365 CRC64;

Query Match 73.3%; Score 487; DB 5; Length 148;
Best Local Similarity 60.8%; Pred. No. 3.6e-47;
Matches 90; Conservative 5; Mismatches 17; Indels 36; Gaps 2;
QY 1 MARKTCGFSKKMLVIAIVMCLCTECYCHCTGAGADCTGCTGAGCNCNPAHTCTDSKN 60
DB 1 MSFKSTFTKMLIIVMCLCNEYNQCTGAGDCTGCTGAGCNCNPAHTCTDSKN 60
QY 61 CVRAATCTGTCNATCTCNSKDCFEAKTCTSTNCYKATCTACTNSTGCGPH 112
DB 61 CVRAATCTGTCNATCTCNSKDCFEAKTCTSTNCYKATCTACTNSTGCGPH 112
QY 88 ---AKTCTDSTNCYKATCTCNSKDCFEAKTCTSTNCYKATCTACTNSTGCGPH 112
DB 121 CYTATCTCNSKATCTACTNSTGCGPH 148
RESULT 10
ID 09NCR4 PRELIMINARY; PRT; 118 AA.
AC 09NCR4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE ANTIFREEZE PROTEIN 12.
GN AFP-12.
OS Dendroidea canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroidea.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;
RT "Isolation and characterization of cDNA clones encoding antifreeze
proteins of the pyrochroid beetle Dendroidea canadensis.";
RL J. Insect Physiol. 46:365-372(2000).
DR EMBL; AF179416; AAF6367.1; -.
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 7.
DR Sequence 118 AA; 12162 MW; 7DCC9542E4BD11B CRC64;
SQ SEQUENCE 118 AA; 12162 MW; 7DCC9542E4BD11B CRC64;
Query Match 58.0%; Score 385; DB 5; Length 118;
Best Local Similarity 61.4%; Pred. No. 7.7e-36;
Matches 70; Conservative 9; Mismatches 21; Indels 14; Gaps 2;
QY 11 KMLVIAIVMCLCTECYCHCTGAGADCTGCTGAGCNCNPAHTCTDSKN 69
DB 4 KALIISVLMCICHEYVAVCTGGPNCNACTTACTCNCNPAHTCTDSKN 63
QY 70 STKCMATCTCNSKDCFEAKTCTSTNCYKATCTACTNSTGCGPH 110
DB 64 STKCMATCTCNSKDCFEAKTCTSTNCYKATCTACTNSTGCGPH 117
RESULT 11
ID 09NCR4 PRELIMINARY; PRT; 122 AA.
AC 09NCR4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE ANTIFREEZE PROTEIN 5 (FRAGMENT).
GN AFP-5.
OS Dendroidea canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroidea.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;

GN AFP-10.
OS Dendroides canadensis.

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Prelygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Cucujiformia; Pyrochroidae; Dendroidea;
 OX NCBI_TaxID=51100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Andorfer C.A., Duman J.G.;
 RT "Isolation and characterization of cDNA clones encoding antifreeze
 RT proteins of the pyrochroid beetle, *Dendroidea canadensis*.";
 RL J. Insect Physiol. 46:365-372(2000).
 DR EMBL; AF179414; AAF6365.1; -.
 DR InterPro; IPR003460; AFP.
 DR Pfam; PF02420; AFP; 8.
 FT NON_TER 1
 SQ SEQUENCE 123 AA; 12882 MW; A5D92CEAD81B4DA5 CRC64;
 Query Match 54.9%; Score 364.5; DB 5; Length 123;
 Best Local Similarity 61.9%; Pred. No. 1.6e-33;
 Matches 70; Conservative 8; Mismatches 20; Indels 15; Gaps 4;
 QY 13 LVIAVIYK-CICTECYC-HCTGGADCTCTDAGCGCNCPCNAHT-CTDSKNCKYKAATCTG 69
 Db 2 LIIGFVLMSCLCHEVYAOECTGSSDCSSCYVACTNCQNCNALPACTDSTNCKYKAFTCTP 61
 QY 70 STKCNARTCTNSKCFE-----AKTCTDSTNCKYKAATCTNSTGCP 110
 Db 62 STKCNAGVTCIDSDCENAEETCTGNTACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 114

Search completed: October 21, 2002, 16:39:38
 Job time : 38.2353 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:03 : Search time 9.35294 Seconds
(without alignments)
285.020 Million cell updates/sec

Title: US-10-032-658-4

Perfect score: 131

Sequence: 1 XCTGXADCTCTACTGCGXCPNA 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 126 | 96.2 | 124 | 20 | AA09281 |
| 2 | 126 | 96.2 | 124 | 20 | AA09282 |
| 3 | 126 | 96.2 | 148 | 20 | AA09280 |
| 4 | 124 | 94.7 | 24 | 20 | AA09276 |
| 5 | 124 | 94.7 | 112 | 20 | AA09278 |
| 6 | 124 | 94.7 | 112 | 20 | AA09279 |
| 7 | 97 | 74.0 | 108 | 18 | AA007693 |
| 8 | 97 | 74.0 | 109 | 18 | AA007694 |
| 9 | 68 | 51.9 | 1679 | 22 | AA007343 |
| 10 | 59 | 45.0 | 233 | 21 | AA074791 |
| 11 | 58 | 44.3 | 24 | 17 | AA098205 |

| | | | | | | |
|----|------|------|------|----|----------|--------------------|
| 12 | 57 | 43.5 | 63 | 22 | ABB66560 | Drosophila melanog |
| 13 | 57 | 43.5 | 119 | 22 | AAG31740 | C glutamicum prote |
| 14 | 56.5 | 43.1 | 524 | 22 | AA007370 | G protein-coupled |
| 15 | 56 | 42.7 | 1081 | 20 | AA024319 | Mouse dephosphoryl |
| 16 | 55.5 | 42.4 | 225 | 21 | AA036084 | Neospora caninum N |
| 17 | 54 | 41.2 | 122 | 21 | AA042599 | Human ORFX ORP2363 |
| 18 | 53 | 40.5 | 1284 | 22 | ABB61690 | Drosophila melanog |
| 19 | 52 | 39.7 | 16 | 20 | AA09275 | Antifreeze protein |
| 20 | 52 | 39.7 | 727 | 11 | AA005533 | Fragment of Heyman |
| 21 | 52 | 39.7 | 3457 | 15 | AA062504 | Large polyprotein |
| 22 | 52 | 39.7 | 3457 | 15 | AA062504 | Polyprotein encode |
| 23 | 51.5 | 39.3 | 86 | 22 | ABG26420 | Novel human diagno |
| 24 | 51.5 | 39.3 | 1798 | 19 | AA050896 | Human laminin B2 c |
| 25 | 51 | 38.9 | 59 | 21 | AA057812 | Trout metallothin |
| 26 | 50.5 | 38.5 | 61 | 19 | AA061601 | Human metallothin |
| 27 | 50.5 | 38.5 | 61 | 21 | AA057822 | Rabbit liver metal |
| 28 | 50.5 | 38.5 | 68 | 12 | AA044774 | Brain-derived grow |
| 29 | 50.5 | 38.5 | 68 | 13 | AA025720 | Nerve nutrient act |
| 30 | 50.5 | 38.5 | 68 | 15 | AA053383 | Polypeptide having |
| 31 | 50.5 | 38.5 | 68 | 22 | AA078444 | Human protein SEQ |
| 32 | 50.5 | 38.5 | 82 | 22 | AA079428 | Human protein SEQ |
| 33 | 50.5 | 38.5 | 215 | 21 | AA056851 | Arabidopsis thalia |
| 34 | 50.5 | 38.5 | 225 | 21 | AA056850 | Arabidopsis thalia |
| 35 | 50.5 | 38.5 | 230 | 21 | AA056849 | Arabidopsis thalia |
| 36 | 50.5 | 38.5 | 420 | 20 | AA013362 | Amino acid sequenc |
| 37 | 50.5 | 38.5 | 420 | 20 | AA005281 | EGF-like homologue |
| 38 | 50.5 | 38.5 | 420 | 21 | AA024396 | Human PRO214 prote |
| 39 | 50.5 | 38.5 | 420 | 21 | AA088569 | Human PRO214 amino |
| 40 | 50.5 | 38.5 | 420 | 22 | AA039899 | Human polypeptide |
| 41 | 50.5 | 38.5 | 420 | 22 | AA012316 | Human PRO214 polyp |
| 42 | 50.5 | 38.5 | 420 | 22 | AA068594 | PRO214. Homo sapi |
| 43 | 50.5 | 38.5 | 420 | 22 | AA080230 | Human PRO214 prote |
| 44 | 50.5 | 38.5 | 420 | 22 | AA048106 | Human TANCO 206 po |
| 45 | 50.5 | 38.5 | 420 | 22 | AA048110 | Mouse TANCO 206 po |

ALIGNMENTS

RESULT 1
AA09281
ID AA09281 standard; Protein; 124 AA.
XX
AC AA09281;
XX
XX 06-JUL-1999 (first entry)
XX
XX
DE YL-3 thermal hysteresis protein.
XX
XX Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX
XX Tenebrio molitor.
OS
XX
XX WO9900493-A1.
PN
XX 07-JAN-1999.
PD
XX
XX 25-JUN-1998; 98WO-CA000618.
PF
XX 26-JUN-1997; 97US-0882907.
PR
XX
XX (TOOH) UNITV QUEENS KINGSTON.
PA
XX Davies PL, Graham LA, Liou Y, Walker VK;
PI WPI; 1999-095739/08.
XX N-PSDB; AAX33468.
DR
XX New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PT quality of foods or biological materials during frozen storage
XX

PS Disclosure; Page 66; 88pp; English.

XX The present invention describes a nucleic acid (I) encoding an antifreeze

CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;

CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c)

CC (1) binds specifically to antibodies raised against the Tenebrio molitor

CC YL1-YL4 antifreeze proteins, (11) is at least 60% identical with YL1-4

CC or (11) has a sequence that includes at least one repeating unit of the

CC consensus sequence CXXSXXCXHX (1). (II), optionally present in an

CC organism, are used to improve preservation of frozen foods (specifically

CC freeze-thaw cycles) or viability of biological materials (e.g. organs,

CC cells or extracts) stored or transported at low temperature, or more

CC generally to depress the freezing point of any aqueous solution,

CC particularly where this is applied to an organism. (I) is used to

CC produce transgenic animals and plants (e.g. grapes, oilseed crops,

CC grains, citrus fruit or sugar cane) having better tolerance to freezing.

CC Fragments of (I) are used as probes to detect or isolate coding

CC sequences, in diagnosis (for determining gene expression) and as

CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to

CC detect (II) is usual immunoassays. (II) from T. molitor have specific

CC activity 100 times greater than that of fish antifreeze proteins. The

CC present sequence represents a thermal hysteresis protein (THP) from the

CC present invention.

XX

SQ Sequence 124 AA;

Query Match 96.2%; Score 126; DB 20; Length 124;

Best Local Similarity 87.0%; Pred. No. 1.4e-07;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 CTGXADCTSCCTACGCGCPCNA 24

||| ||||| ||||| |||||

Db 30 CTGADCTSCCTACGCGCPCNA 52

RESULT 2

AAV09282

ID AAV09282 standard; Protein; 124 AA.

XX

AC AAV09282;

XX

DT 06-JUL-1999 (first entry)

XX

DE 5-15 thermal hysteresis protein.

XX

KW Antifreeze; thermal hysteresis protein; THP; mealworm beetle;

KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.

XX

OS Tenebrio molitor.

XX

PN WO9900493-A1.

XX

PD 07-JAN-1999.

XX

PF 25-JUN-1998; 98WO-CA00618.

XX

PR 26-JUN-1997; 97US-0882907.

XX

PA (TOOH) UNIV QUEBENS KINGSTON.

XX

PI Davies PL, Graham LA, Liou Y, Walker VK;

XX

DR WPI: 1999-095739/08.

DR N-PSDB; AAX33469.

XX

PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -

PT used to improve low temperature tolerance of fish, plants etc., and

PT quality of foods or biological materials during frozen storage

XX

PS Disclosure; Page 67-68; 88pp; English.

XX

CC The present invention describes a nucleic acid (I) encoding an antifreeze

CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;

CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c)

CC (1) binds specifically to antibodies raised against the Tenebrio molitor

CC YL1-YL4 antifreeze proteins, (11) is at least 60% identical with YL1-4

CC or (11) has a sequence that includes at least one repeating unit of the

CC consensus sequence CXXSXXCXHX (1). (II), optionally present in an

CC organism, are used to improve preservation of frozen foods (specifically

CC freeze-thaw cycles) or viability of biological materials (e.g. organs,

CC cells or extracts) stored or transported at low temperature, or more

CC generally to depress the freezing point of any aqueous solution,

CC particularly where this is applied to an organism. (I) is used to

CC produce transgenic animals and plants (e.g. grapes, oilseed crops,

CC grains, citrus fruit or sugar cane) having better tolerance to freezing.

CC Fragments of (I) are used as probes to detect or isolate coding

CC sequences, in diagnosis (for determining gene expression) and as

CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to

CC detect (II) is usual immunoassays. (II) from T. molitor have specific

CC activity 100 times greater than that of fish antifreeze proteins. The

CC present sequence represents a thermal hysteresis protein (THP) from the

CC present invention.

XX

SQ Sequence 124 AA;

Query Match 96.2%; Score 126; DB 20; Length 124;

Best Local Similarity 87.0%; Pred. No. 1.4e-07;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 CTGXADCTSCCTACGCGCPCNA 24

||| ||||| ||||| |||||

Db 30 CTGADCTSCCTACGCGCPCNA 52

RESULT 3

AAV09280

ID AAV09280 standard; Protein; 148 AA.

XX

AC AAV09280;

XX

DT 06-JUL-1999 (first entry)

XX

DE YL-4 thermal hysteresis protein.

XX

KW Antifreeze; thermal hysteresis protein; THP; mealworm beetle;

KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.

XX

OS Tenebrio molitor.

XX

PN WO9900493-A1.

XX

PD 07-JAN-1999.

XX

PF 25-JUN-1998; 98WO-CA00618.

XX

PR 26-JUN-1997; 97US-0882907.

XX

PA (TOOH) UNIV QUEBENS KINGSTON.

XX

PI Davies PL, Graham LA, Liou Y, Walker VK;

XX

DR WPI: 1999-095739/08.

DR N-PSDB; AAX33467.

XX

PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -

PT used to improve low temperature tolerance of fish, plants etc., and

PT quality of foods or biological materials during frozen storage

XX

PS Disclosure; Page 64; 88pp; English.

XX

CC The present invention describes a nucleic acid (I) encoding an antifreeze

CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;

CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c)

CC (1) binds specifically to antibodies raised against the Tenebrio molitor

CC YL1-YL4 antifreeze proteins, (II) is at least 60% identical with YL1-4
CC or (III) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXSKXXCXXXT (I). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically
CC to reduce formation of ice crystals following exposure to repeated
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more
CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (II) is usual immunoassays. (II) from T. mollitor have specific
CC activity 100 times greater than that of fish antifreeze proteins. The
CC present sequence represents a thermal hysteresis protein (THP) from the
CC present invention.
CC
CC
SQ Sequence 148 AA;

Query Match 96.2%; Score 126; DB 20; Length 148;
Best Local Similarity 87.0%; Pred. No. 1.7e-07;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXCPNA 24
||| ||||| ||||| |||||
Db 30 CTGAADCTCTACTGCGCNCNA 52

RESULT 4
AA09276
ID AAY09276 standard; peptide: 24 AA.

AC AAY09276;

DT 06-JUL-1999 (first entry)

DE Antifreeze protein amino acid sequence SEQ ID NO:4.

XX Antifreeze; thermal hysteresis protein; THP; mealworm beetle;

KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.

XX Synthetic.

OS Tenebrio molitor.

XX Key Location/Qualifiers

FT Misc-difference 1..24 /note= "X is unspecified"

PN WO9900493-A1.

PD 07-JAN-1999.

PF 25-JUN-1998; 98WO-CA00618.

PR 26-JUN-1997; 97US-0882907.

XX (TOOH) UNIV QUEBENS KINGSTON.

PI Davies PL, Graham LA, Liou Y, Walker VK;

DR WPI: 1999-095739/08.

XX New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PR quality of foods or biological materials during frozen storage

XX Claim 14; Page 56; 88pp; English.

CC The present invention describes a nucleic acid (I) encoding an antifreeze
CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/mL; (c)
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/mL; (c)

CC (I) binds specifically to antibodies raised against the Tenebrio molitor
CC YL1-YL4 antifreeze proteins, (II) is at least 60% identical with YL1-4
CC or (III) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXSKXXCXXXT (I). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically
CC to reduce formation of ice crystals following exposure to repeated
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more
CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (II) is usual immunoassays. (II) from T. mollitor have specific
CC activity 100 times greater than that of fish antifreeze proteins.
CC
CC
SQ Sequence 24 AA;

Query Match 94.7%; Score 124; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXCPNA 24
||| ||||| ||||| |||||
Db 2 CTGXADCTCTXACTGCGXCPNA 24

RESULT 5
AA09278
ID AAY09278 standard; Protein; 112 AA.

AC AAY09278;

DT 06-JUL-1999 (first entry)

DE YL-1 thermal hysteresis protein.

XX Antifreeze; thermal hysteresis protein; THP; mealworm beetle;

KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.

XX Tenebrio molitor.

OS Tenebrio molitor.

XX Key Location/Qualifiers

FT Misc-difference 1..24 /note= "X is unspecified"

PN WO9900493-A1.

PD 07-JAN-1999.

PF 25-JUN-1998; 98WO-CA00618.

PR 26-JUN-1997; 97US-0882907.

XX (TOOH) UNIV QUEBENS KINGSTON.

PI Davies PL, Graham LA, Liou Y, Walker VK;

DR WPI: 1999-095739/08.

XX New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PR quality of foods or biological materials during frozen storage

XX Disclosure; Page 60; 88pp; English.

CC The present invention describes a nucleic acid (I) encoding an antifreeze
CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/mL; (c)
CC (I) binds specifically to antibodies raised against the Tenebrio molitor
CC YL1-YL4 antifreeze proteins, (II) is at least 60% identical with YL1-4
CC or (III) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXSKXXCXXXT (I). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically

cells or extracts) stored or transported at low temperature, or more

PR 07-JUN-1995; 95US-0485359.

```

XX (UNOT ) UNIV NOTRE DAME DU LAC.
PA
XX Duman JG;
PI
XX WPI: 1997-052352/05.
DR N-PSDB; AAT47150.
XX
PT DNA encoding Dendroides sp. thermal hysteresis protein - produces
PT protein having antifreeze properties, useful to protect plant cells,
PT seeds or plants from frost damage
XX
PS Disclosure: Page 31-32; 46pp; English.
XX
CC The present sequence represents a novel peptide which has antifreeze
CC properties. The nucleic acid sequence was derived from Dendroides
CC canadensis, and this protein is an example of a thermal-hysteresis
CC protein. The mature protein sequence contains a 13 amino acid sequence
CC and a 12 amino acid sequence that are repeated three and four times
CC respectively (see features table). The nucleic acid sequence can
CC be used to transform a plant cell for the heterologous production of
CC Dendroides thermal hysteresis protein (THP), i.e. to protect it from
CC frost damage. The DNA and polynucleotide sequences may be used as probes
CC for the isolation of THP coding sequences, from other organisms.
XX
SQ Sequence 108 AA;

Query Match 74.0%; Score 97; DB 18; Length 108;
Best Local Similarity 65.2%; Pred. No. 0.00026;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGXADCTCTCTACTGCGXCPNA 24
|||:|||||:|||||
DB 27 CTGGSDCRSCVTSCDQNCNCPNA 49

RESULT 8
AAW07694
ID AAW07694 standard; Protein: 109 AA.
XX
XX AAW07694;
AC
XX 09-SEP-1997 (first entry)
DT
XX
DE Dendroides canadensis thermal-hysteresis protein.
XX
XX THP; antifreeze; transformation; plant.
KW
XX Dendroides canadensis.
OS
XX
FH Key Location/Qualifiers
FT 1..19 /label= Signal
FT 20..108 /label= Signal
FT Protein /product= Thermal-hysteresis_protein
FT Region 20..32
FT /label= Repeat_A
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region 33..45
FT /label= Repeat_B
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region 46..58
FT /label= Repeat_C
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region 59..70
FT /label= Repeat_D
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region 7182..82
FT /label= Repeat_E
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region 83..94
FT /label= Repeat_F
FT /note= "See AAW07697 for 12 amino acid consensus

```

```

FT sequence" Region 95..104
FT /label= Repeat_G
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence"
XX
XX WO9640973-A1.
XX
XX 19-DEC-1996.
XX
XX
XX 05-JUN-1996; 96WO-US08815.
XX
XX 07-JUN-1995; 95US-0485359.
XX 08-DEC-1995; 95US-0569594.
XX
XX (UNOT ) UNIV NOTRE DAME DU LAC.
PA
XX Duman JG;
PI
XX WPI: 1997-052352/05.
DR N-PSDB; AAT47151.
XX
PT DNA encoding Dendroides sp. thermal hysteresis protein - produces
PT protein having antifreeze properties, useful to protect plant cells,
PT seeds or plants from frost damage
XX
XX
PS Disclosure: Page 33; 46pp; English.
XX
CC The present sequence represents a novel peptide which has antifreeze
CC properties. The nucleic acid sequence was derived from Dendroides
CC canadensis, and this protein is an example of a thermal-hysteresis
CC protein. The mature protein sequence contains a 13 amino acid sequence
CC and a 12 amino acid sequence that are repeated three and four times
CC respectively (see features table). The nucleic acid sequence can
CC be used to transform a plant cell for the heterologous production of
CC Dendroides thermal hysteresis protein (THP), i.e. to protect it from
CC frost damage. The DNA and polynucleotide sequences may be used as probes
CC for the isolation of THP coding sequences, from other organisms.
XX
SQ Sequence 109 AA;

Query Match 74.0%; Score 97; DB 18; Length 109;
Best Local Similarity 65.2%; Pred. No. 0.00027;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGXADCTCTCTACTGCGXCPNA 24
|||:|||||:|||||
DB 27 CTGGSDCRSCVTSCDQNCNCPNA 49

RESULT 9
AAU07343
ID AAU07343 standard; Protein: 1679 AA.
XX
XX AAU07343;
AC
XX 04-DEC-2001 (first entry)
DT
XX
DE 1-aminocyclopropane carboxylate (ACPC) synthase #12.
XX
XX 1-aminocyclopropane carboxylate synthase; ACPC synthase; brain trauma;
XX excitatory neurotransmission; neurodegeneration; stroke; nerve damage;
XX neurodegenerative disease; Alzheimer's disease; depression; epilepsy;
XX alcohol abuse; cognitive function; memory; learning impairment; human.
XX
XX Homo sapiens.
XX
XX WO200168879-A2.
XX
XX 20-SEP-2001.
XX
XX 14-MAR-2001; 2001WO-EP02857.
XX
XX 14-MAR-2000; 2000US-0189086.
XX

```

```

PR 05-APR-2000; 2000US-0194702.
PA (FARB ) BAYER AG.
PI
PI Ramakrishnan S;
PI
PI WPI; 2001-550286/61.
DR
DR
PT Isolated polynucleotide encoding a human
PT 1-aminocyclopropane-carboxylate (ACPC) synthase, useful for treating
PT brain trauma and neurodegenerative disease (e.g. Alzheimer's disease,
PT depression, epilepsy) -
PS
PS Claim 1; Page 231-237; 242pp; English.
XX
XX The invention relates to reagents and methods for regulating excitatory
XX neurotransmission, and to prevent neurodegeneration. The method involves
XX the use of an expression vector or a reagent that modulates the activity
XX of a 1-aminocyclopropane-carboxylate (ACPC) synthase polypeptide. The
XX reagent is useful for modulating the activity of an ACPC synthase in a
XX disease such as stroke, a nerve damage or a neurodegenerative disease.
XX The ACPC synthase polypeptide, polynucleotides and modulators are also
XX useful for creating brain trauma and neurodegenerative disease (e.g.
XX Alzheimer's disease, depression, epilepsy). The ACPC synthase modulators
XX are also useful for treating alcohol abuse and improve cognitive function
XX and memory of patients with learning impairment. The present sequence
XX represents the amino acid sequence of human 1-aminocyclopropane-
XX carboxylate (ACPC) synthase #12, used in the method of the invention.
SQ
SQ Sequence 1679 AA;
Query Match 51.9%; Score 68; DB 22; Length 1679;
Best Local Similarity 50.0%; Pred. No. 5.1;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
OY 2 CTGXADCTCTACTGCGXC 21
DB 1112 CCGTGCCTTCCTCTGCGCGC 1131
RESULT 10
AAV74791
ID AAV74791 standard; Protein; 233 AA.
XX
XX AAV74791;
XX
XX 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 263 protein sequence SEQ ID NO:1056.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
XX
XX WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.

```

```

XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masiagnani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR
DR N-PSDB; AA253553.
XX
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
PS
PS Claim 2; Page 606; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
XX represent novel Neisseria meningitidis and N. gonorrhoea polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
SQ
SQ Sequence 233 AA;
Query Match 45.0%; Score 59; DB 21; Length 233;
Best Local Similarity 45.8%; Pred. No. 11;
Matches 11; Conservative 0; Mismatches 9; Indels 4; Gaps 1;
OY 2 CTGXADCTCTCT---XACTGCGXC 21
DB 159 CTGGCAACCTGTGCACTACGTC 182
RESULT 11
AAR98205
ID AAR98205 standard; Protein; 24 AA.
XX
XX AAR98205;
XX
XX 30-DEC-1996 (first entry)
DE Primer for amplifying MSH/MOMULV chimeric protein coding sequence.
XX
XX Moloney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;
KW 10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;
KW polytropic MX27 provirus; targeted drug delivery; gene therapy;
KW single chain antibody; envelope protein; ss.
XX
XX Synthetic.
XX
XX WO9630504-A1.
XX
XX 03-OCT-1996.
XX
XX 22-MAR-1996; 96WO-US03908.
XX
XX 24-MAR-1995; 95US-0409648.
XX
XX (GENE-) GENETIC THERAPY INC.
XX (UTSC-) UNIV SOUTHERN CALIFORNIA.
XX
XX Anderson W, Chiang YL, Januszewski M, Mackrell AJ;
PI Zhao Y;
XX
XX WPI; 1996-455352/45.
XX
XX Cell-targeted retroviral vector particles - having envelope protein
PT modified with targeting polypeptide

```

XX Example 2; Page 36; 73pp; English.
PS
XX
CC Cell targeted retroviral vector particles can be used in gene
CC therapy to deliver a heterologous gene to a target cell for
CC expression of a heterologous polypeptide in that cell. The cell
CC targeted retroviral vector particles comprise an envelope protein
CC which is modified to contain a targeting polypeptide (a single chain
CC antibody), or in the case of moloney murine leukaemia virus
CC (MOMuLV), alpha melanotropin-stimulating hormone (MSH). Four primers
CC (AAT33931, AAR98205 and AAT33932, AAR98206) were used to screen and
CC sequence for synthetic MSH/MOMuLV chimeric protein coding sequences.
CC Two other oligonucleotides used to substitute sequences in MOMuLV for
CC MSH sequences are described in AAR98207 and AAR98208.
XX
SQ Sequence 24 AA;
Query Match 44.3%; Score 58; DB 17; Length 24;
Best Local Similarity 50.0%; Pred. No. 2.2;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
QY 4 GXAAGCTGCTACGCGXC 21
Db 5 GCAACTCTGTGCAACACAC 22
RESULT 12
ABB66560
ID ABB66560 standard; Protein; 63 AA.
XX
AC ABB66560;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 26472.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li FWD, Myers EW;
PI
DR WPI: 2001-656860/75.
DR N-PSDB; ABL10663.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 26472; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57731-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 63 AA;
Query Match 43.5%; Score 57; DB 22; Length 63;
Best Local Similarity 45.0%; Pred. No. 6.2;
Matches 9; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2 CCGXACTGCTACGCGXC 21
Db 20 CGGCGPCGGCGCGCGCGC 39
RESULT 13
AAG91740
ID AAG91740 standard; Protein; 119 AA.
XX
AC AAG91740;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 5494.
XX
KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOWA) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
PI
DR WPI: 2001-376931/40.
DR N-PSDB; AAH66959.
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17; SEQ ID NO: 5494; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 119 AA;
Query Match 43.5%; Score 57; DB 22; Length 119;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 10; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

QY 9 TSCYACTGCG-----XCP 22
|||:|||||
Db 93 TSCSLCTGCGCKTVAVTCTP 112

RESULT 14
AAU07370

ID AAU07370 standard; Protein: 524 AA.

XX AAU07370;

AC XX 18-DEC-2001 (first entry)

DE G protein-coupled receptor.

XX Human: mental disorder; thyroid disease; renal failure; anorexia;
KW inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;
KW autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;
KW depression; Parkinson's disease; Alzheimer's disease; viral infection;
KW Huntington's disease; human immunodeficiency virus; type 2 diabetes;
KW anorexia; hypotension; hypertension; thrombosis; myocardial infarction;
KW atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;
KW nPCR.

XX Homo sapiens.

XX WO200162924-A2.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-US05989.

XX 24-FEB-2000; 2000US-0184602.
PR 24-FEB-2000; 2000US-0184604.
PR 24-FEB-2000; 2000US-0184606.
PR 24-FEB-2000; 2000US-0184689.
PR 24-FEB-2000; 2000US-0184690.
PR 24-FEB-2000; 2000US-0184710.
PR 24-FEB-2000; 2000US-0184712.
PR 24-FEB-2000; 2000US-0184715.
PR 24-FEB-2000; 2000US-0184716.
PR 24-FEB-2000; 2000US-0184725.
PR 24-FEB-2000; 2000US-0184822.
XX (PHMA) PHARMACIA & UPJOHN CO.

XX Vogel G, Wood LS, Parodi LA, Lind P;
PI WPI: 2001-570632/64.
XX Novel nucleic acid and encoded nPCR-X, used to screen for compounds
PT for use in the treatment of mental disorders, such as Alzheimer's
PT disease, or Parkinson's disease -
XX Claim 31; Page 178-179; 263pp; English.

XX The invention relates to novel isolated human G protein-coupled
CC receptors (nPCR-X). The nPCR-X can be used for screening compounds
CC which can be used to treat mental disorders, thyroid disease, renal
CC failure, inflammatory conditions such as Crohn's disease, rheumatoid
CC arthritis, autoimmune disorders, schizophrenia, migraine, stroke,
CC dementia, depression, Parkinson's disease, Alzheimer's disease, and
CC Huntington's disease. They may also be used for treating viral infections
CC such as human immunodeficiency virus (HIV), type 2 diabetes, obesity,
CC anorexia, hypotension, hypertension, thrombosis, myocardial infarction,
CC atherosclerosis, cancer, and sexual dysfunction. AAU25617-AAU25726
CC represent the amino acid sequences of novel human G protein-coupled
CC receptors, nPCR-2031 to nPCR-2140 respectively, as described in the
CC invention.

SQ Sequence 524 AA;

Query Match 43.1%; Score 56.5; DB 22; Length 524;

Best Local Similarity 44.8%; Pred. NO. 40;
Matches 13; Conservative 1; Mismatches 6; Indels 9; Gaps 2;

QY 2 CTGKAD-----CTGCTXA---CTGCGXC 21

Db 172 CTGCATATCAGACTATTTAAAGCAGCGCC 200

RESULT 15

ID AAU24319 standard; Protein: 1081 AA.

XX AAU24319;

AC XX 16-SEP-1999 (first entry)

DE Mouse dephosphorylase inhibiting p91-like protein #2.

XX Dephosphorylase inhibiting protein; p91; tyrosine phosphatase SHP-1;
KW SHP-2; inositol-5-phosphate SHP; phosphorylating tyrosine;
KW immunoreceptor; immunomodulatory agent.
KW Mus sp.

XX JP11169184-A.
XX 29-JUN-1999.

PF 12-DEC-1997; 97JP-0362285.

XX 12-DEC-1997; 97JP-0362285.

PA (UYOK-) UNIV OKAYAMA.

XX WPI: 1999-422622/36.
DR N-PSDB: AAX88976.

PT New peptide - useful for inhibiting dephosphorylase

XX Claim 2; Page 15-17; 30pp; Japanese.

XX The present invention describes new proteins for inhibiting
CC dephosphorylase. The proteins can be combined with tyrosine phosphatase
CC SHP-1, SHP-2 or inositol-5-phosphatase SHP by phosphorylating tyrosine.
CC The present invention also describes an immunoreceptor comprising one
CC the above proteins, and DNA coding the above proteins. The new proteins
CC can be used as an immunomodulatory agent. The present sequence
CC represents a protein from the present invention.

SQ Sequence 1081 AA;

Query Match 42.7%; Score 56; DB 20; Length 1081;
Best Local Similarity 47.8%; Pred. NO. 82;
Matches 11; Conservative 1; Mismatches 7; Indels 4; Gaps 2;

QY 2 CTGXADCTCTYXACTGCGCXPA 24

Db 744 CAG---CAACC-CTGCGGACTAA 762

Search completed: October 21, 2002, 16:38:22
Job time : 10.3529 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:37:29 : Search time 3.70588 Seconds
(without alignments)
158.185 Million cell updates/sec

Title: US-10-032-658-4

Perfect score: 131

Sequence: 1 XCTGXADCTSCYACTGCGXCPNA 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 segs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 97 | 74.0 | 108 | 1 | US-08-485-359-2 |
| 2 | 97 | 74.0 | 108 | 1 | US-08-569-594-2 |
| 3 | 97 | 74.0 | 108 | 5 | PCT-US96-08815-2 |
| 4 | 97 | 74.0 | 109 | 1 | US-08-485-359-4 |
| 5 | 97 | 74.0 | 109 | 1 | US-08-569-594-4 |
| 6 | 97 | 74.0 | 109 | 5 | PCT-US96-08815-4 |
| 7 | 97 | 54.2 | 1345 | 2 | US-08-977-767-3 |
| 8 | 70.5 | 53.8 | 120 | 3 | US-08-508-761B-22 |
| 9 | 68.5 | 52.3 | 45 | 4 | US-08-300-230-14 |
| 10 | 66.5 | 50.8 | 801 | 1 | US-07-906-349A-6 |
| 11 | 66.5 | 50.8 | 1400 | 4 | US-08-630-915A-37 |
| 12 | 65.5 | 50.0 | 1417 | 4 | US-08-900-230-3 |
| 13 | 63.5 | 48.5 | 57 | 1 | US-07-609-716-56 |
| 14 | 61 | 46.6 | 47 | 3 | US-08-482-085B-91 |
| 15 | 60 | 45.8 | 45 | 4 | US-08-900-230-8 |
| 16 | 60 | 45.4 | 45 | 4 | US-08-900-230-11 |
| 17 | 59.5 | 45.4 | 45 | 4 | US-08-900-230-45 |
| 18 | 59 | 45.0 | 3788 | 4 | US-09-336-447A-76 |
| 19 | 58.5 | 44.7 | 50 | 4 | US-08-900-230-58 |
| 20 | 58.5 | 44.7 | 54 | 1 | US-08-279-058B-24 |
| 21 | 58.5 | 44.7 | 102 | 3 | US-08-974-022-53 |
| 22 | 58.5 | 44.7 | 102 | 4 | US-08-795-445A-53 |
| 23 | 58.5 | 44.7 | 102 | 4 | US-08-795-447A-53 |
| 24 | 58.5 | 44.7 | 102 | 4 | US-08-974-186-53 |
| 25 | 58.5 | 44.7 | 102 | 4 | US-08-795-446B-53 |
| 26 | 58.5 | 44.7 | 102 | 4 | US-08-990-823-112 |
| 27 | 58.5 | 44.7 | 143 | 4 | US-08-990-823-112 |

| | | | | | | |
|----|------|------|-----|---|-------------------|-------------------|
| 28 | 58 | 44.3 | 45 | 4 | US-08-900-230-16 | Sequence 16, App1 |
| 29 | 57.5 | 43.9 | 341 | 2 | US-08-209-521-11 | Sequence 11, App1 |
| 30 | 57 | 43.5 | 52 | 1 | US-07-609-716-68 | Sequence 68, App1 |
| 31 | 57 | 43.5 | 55 | 3 | US-08-476-509B-46 | Sequence 46, App1 |
| 32 | 55.5 | 42.4 | 46 | 4 | US-08-900-230-40 | Sequence 40, App1 |
| 33 | 54 | 41.2 | 45 | 4 | US-08-900-230-43 | Sequence 43, App1 |
| 34 | 54 | 41.2 | 45 | 4 | US-08-900-230-44 | Sequence 44, App1 |
| 35 | 54 | 41.2 | 48 | 5 | PCT-US96-01720-8 | Sequence 8, App1 |
| 36 | 53.5 | 40.8 | 33 | 4 | US-09-047-288-4 | Sequence 4, App1 |
| 37 | 53.5 | 40.8 | 45 | 4 | US-08-900-230-52 | Sequence 52, App1 |
| 38 | 53.5 | 40.8 | 47 | 1 | US-08-451-947-95 | Sequence 95, App1 |
| 39 | 53.5 | 40.8 | 47 | 1 | US-08-451-947-96 | Sequence 96, App1 |
| 40 | 53.5 | 40.8 | 47 | 2 | US-08-424-826A-95 | Sequence 95, App1 |
| 41 | 53.5 | 40.8 | 47 | 2 | US-08-424-826A-96 | Sequence 96, App1 |
| 42 | 53.5 | 40.8 | 47 | 3 | US-08-928-694-95 | Sequence 95, App1 |
| 43 | 53.5 | 40.8 | 47 | 3 | US-08-928-694-96 | Sequence 96, App1 |
| 44 | 53.5 | 40.8 | 47 | 5 | PCT-US91-06950-95 | Sequence 95, App1 |
| 45 | 53.5 | 40.8 | 47 | 5 | PCT-US91-06950-96 | Sequence 96, App1 |

ALIGNMENTS

```
RESULT 1
: Sequence 2, Application US/08485359
: Patent No. 5627051
: GENERAL INFORMATION:
: APPLICANT: Duman, John G.
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
: DENDROIDES CANADENSIS ANTIFREEZE PROTEINS
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Barnes & Thornburg
: STREET: 11 South Meridian
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: USA
: ZIP: 46204
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/485,359
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Lammert, Steven R.
: REGISTRATION NUMBER: 27653
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317) 231-7258
: TELEFAX: (317) 231-7433
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 108 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: YES
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Dendroides canadensis
: US-08-485-359-2
:
: Query Match 74.0% Score 97; DB 1; Length 108;
: Best local similarity 65.2%; Pred. No. 2.3e-05;
: Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
:
: QY 2 CTGXADCTSCYACTGCGXCPNA 24
: ||| :|| ||| :||| |||||
```

Db 27 CTGSDCRSCTVSCDCCNCPNA 49

RESULT 2

US-08-569-594-2

Sequence 2, Application US/08569594

Patent No. 5633451

GENERAL INFORMATION:

APPLICANT: Duman, John G.

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING DENDROIDES CANADENSIS ANTIFREEZE PROTEINS

TITLE OF INVENTION: DENDROIDES CANADENSIS ANTIFREEZE PROTEINS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Barnes & Thornburg

STREET: 11 South Meridian

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46204

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/569,594

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Lammert, Steven R.

REGISTRATION NUMBER: 27653

REFERENCE/DOCKET NUMBER: 835910-25377

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 231-7258

TELEFAX: (317) 231-7433

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 108 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Dendroides canadensis

US-08-569-594-2

Query Match 74.0%; Score 97; DB 1; Length 108;

Best Local Similarity 65.2%; Pred. No. 2.3e-05;

Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXCPNA 24

||| :||| ||| :||| |||

Db 27 CTGSDCRSCTVSCDCCNCPNA 49

RESULT 3

PCT-US96-08815-2

Sequence 2, Application PC/TUS9608815

GENERAL INFORMATION:

APPLICANT: Duman, John G.

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING DENDROIDES ANTIFREEZE PROTEINS

TITLE OF INVENTION: DENDROIDES ANTIFREEZE PROTEINS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Barnes & Thornburg

STREET: 11 South Meridian

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46204

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/08815

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Lammert, Steven R.

REGISTRATION NUMBER: 27653

REFERENCE/DOCKET NUMBER: 835910-27026

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 231-7258

TELEFAX: (317) 231-7433

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 108 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Dendroides canadensis

PCT-US96-08815-2

Query Match 74.0%; Score 97; DB 5; Length 108;

Best Local Similarity 65.2%; Pred. No. 2.3e-05;

Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXCPNA 24

||| :||| ||| :||| |||

Db 27 CTGSDCRSCTVSCDCCNCPNA 49

RESULT 4

US-08-485-359-4

Sequence 4, Application US/08485359

Patent No. 5627051

GENERAL INFORMATION:

APPLICANT: Duman, John G.

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING DENDROIDES CANADENSIS ANTIFREEZE PROTEINS

TITLE OF INVENTION: DENDROIDES CANADENSIS ANTIFREEZE PROTEINS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Barnes & Thornburg

STREET: 11 South Meridian

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46204

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/485,359

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lammert, Steven R.

REGISTRATION NUMBER: 27653

REFERENCE/DOCKET NUMBER: 835910-25377

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 231-7258

TELEFAX: (317) 231-7433

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 109 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Dendroides canadensis
US-08-485-359-4

Query Match 74.0%; Score 97; DB 1; Length 109;
Best Local Similarity 65.2%; Pred. No. 2.3e-05;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGAXDCTCTACTGCGXCPCNA 24
DB 27 CTGGSDCRSTVCTDQNCPCNA 49

RESULT 5
US-08-569-594-4
Sequence 4, Application US/08569594
Patent No. 5633451
GENERAL INFORMATION:

APPLICANT: Duman, John G.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
DENDROIDES CANADENSIS ANTIFREEZE PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Barnes & Thornburg
STREET: 11 South Meridian
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,594
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Lammett, Steven R.
REGISTRATION NUMBER: 27653
REFERENCE/DOCKET NUMBER: 835910-25377
TELEPHONE: (317) 231-7258
TELEFAX: (317) 231-7433
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Dendroides canadensis
US-08-569-594-4

Query Match 74.0%; Score 97; DB 1; Length 109;
Best Local Similarity 65.2%; Pred. No. 2.3e-05;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGAXDCTCTACTGCGXCPCNA 24
DB 27 CTGGSDCRSTVCTDQNCPCNA 49

RESULT 6
PCT-US96-08815-4
Sequence 4, Application PC/TUS9608815
GENERAL INFORMATION:
APPLICANT: Duman, John G.

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
DENDROIDES ANTIFREEZE PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Barnes & Thornburg
STREET: 11 South Meridian
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08815
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lammett, Steven R.
REGISTRATION NUMBER: 27653
REFERENCE/DOCKET NUMBER: 835910-27026
TELEPHONE: (317) 231-7258
TELEFAX: (317) 231-7433
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Dendroides canadensis
PCT-US96-08815-4

Query Match 74.0%; Score 97; DB 5; Length 109;
Best Local Similarity 65.2%; Pred. No. 2.3e-05;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGAXDCTCTACTGCGXCPCNA 24
DB 27 CTGGSDCRSTVCTDQNCPCNA 49

RESULT 7
US-08-977-767-3
Sequence 3, Application US/08977767
Patent No. 5972684
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Yue, Henry
APPLICANT: Greenwald, Sara
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,767
FILING DATE: Herewith

```

: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0423 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1345 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: GenBank
: CLONE: 1532042
US-08-977-767-3

Query Match          54.2%; Score 71; DB 2; Length 1345;
Best Local Similarity 59.1%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

QY 2 CTGXADCTSC--TXACTGCGXC 21
Db 1024 CTCGAGCTGCACTCCCTGCGAC 1045

RESULT 8
US-08-508-761B-22
: Sequence 22 Application US/08508761B
: Patent No. 6027920
: GENERAL INFORMATION:
: APPLICANT: Jolliff, Gwennael
: APPLICANT: Guyonvarch, Armel
: APPLICANT: Purification, Relano
: APPLICANT: Duchiron, Francis
: APPLICANT: Renaud, Michel
: TITLE OF INVENTION: System for Protein Expression and
: TITLE OF INVENTION: Secretion Especially in Corynebacteria
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: Jacobson, Price, Holman & Stern, PLLC
: STREET: 400 Seventh St. N.W.
: CITY: Washington D.C.
: COUNTRY: U.S.A.
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/508,761B
: FILING DATE: 31-JUL-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91/09652
: FILING DATE: 29-JUL-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91/09870
: FILING DATE: 02-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Player, William E.
: REGISTRATION NUMBER: 31,409
: REFERENCE/DOCKET NUMBER: P58525NA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 638-6666
: TELEFAX: (202) 393-5350
: INFORMATION FOR SEQ ID NO: 22:
```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 120 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: ORIGINAL SOURCE:
: ORGANISM: synthetic
US-08-508-761B-22

Query Match          53.8%; Score 70.5; DB 3; Length 120;
Best Local Similarity 60.0%; Pred. No. 0.036;
Matches 12; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 2 CTGXADCTSCCTGCTGCGXC 21
Db 30 CTGTGCTGTC-GCCTGCGCC 48

RESULT 9
US-08-900-230-14
: Sequence 14 Application US/08900230
: Patent No. 6329197
: GENERAL INFORMATION:
: APPLICANT: Bard, Jonathan A.
: TITLE OF INVENTION: DNA ENCODING GALANN GALR3 RECEPTORS AND
: TITLE OF INVENTION: USES THEREOF
: NUMBER OF SEQUENCES: 59
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: Cooper & Dunham LLP
: STREET: 1185 Avenue of The Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 11036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/900,230
: FILING DATE: 23-JUL-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 52241-C/JPM/ADM
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-278-0400
: TELEFAX: 212-391-0525
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 45 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: MOLECULE TYPE:
: TOPOLOGY: linear
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
US-08-900-230-14

Query Match          52.3%; Score 68.5; DB 4; Length 45;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 1; Mismatches 6; Indels 9; Gaps 1;

QY 2 CTGXADCTSC-----CTXACTGCGXC 21
Db 13 CTCGACCTGCTGCTTACGCTACCTGCTGC 41

RESULT 10
US-07-906-349A-6
```

Sequence 6, Application US/07906349A
Patent No. 5434064
GENERAL INFORMATION:
APPLICANT: Schlessinger, Joseph
APPLICANT: Skolnik, Edward Y.
APPLICANT: Margolis, Benjamin L.
TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASES AND
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/906,349A
FILING DATE: 30-JUN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/643,237
FILING DATE: 18-JAN-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-906-349A-6

Query Match 50.8%; Score 66.5; DB 1; Length 801;
Best Local Similarity 55.0%; Pred. No. 0.55;
Matches 11; Conservative 1; Mismatches 5; Indels 3; Gaps 1;
OY 2 CTGXADCTCTXACTGCGXC 21
||| | :|| || || |
Db 239 CTG---CCTCTGACAGCAGC 255

RESULT 11
US-08-630-915A-37
Sequence 37, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLKES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-37

Query Match 50.8%; Score 66.5; DB 4; Length 1400;
Best Local Similarity 46.7%; Pred. No. 0.88;
Matches 14; Conservative 1; Mismatches 8; Indels 7; Gaps 2;
OY 2 CTGXADCTCTXACTGCGXC 24
||| ||: | |||| | | |
Db 882 CTGCTCTTCAGAGAGTCTGCTAGTCTTA 911

RESULT 12
US-08-900-230-3
Sequence 3, Application US/08900230
Patent No. 6329197
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A.
TITLE OF INVENTION: DNA ENCODING GALANN GAIK3 RECEPTORS AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 11036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,230
FILING DATE: 23-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52241-C/JPW/ADM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO
HYPOTHETICAL: NO

ANTI-SENSE: NO
US-08-900-230-3
Query Match 50.0%; Score 65.5; DB 4; Length 1417;
Best Local Similarity 52.0%; Pred. No. 1.2;
Matches 13; Conservative 1; Mismatches 6; Indels 5; Gaps 2;
QY 2 CTGACCTCTTACGCTGCTGTC 525
DB 501 CAGCAGCTTACGCTGCTGCTGTC 525
RESULT 13
US-07-609-716-56
; Sequence 56, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-07-609-716-56
Query Match 48.5%; Score 63.5; DB 1; Length 57;
Best Local Similarity 55.0%; Pred. No. 0.13;
Matches 11; Conservative 1; Mismatches 7; Indels 1; Gaps 1;
QY 2 CTGACCTCTTACGCTGCTGTC 21
DB 28 CCGGCGCTTACCTACTG-GAC 46
RESULT 14
US-08-482-085B-91
; Sequence 91, Application US/08482085B
; Patent No. 6018030
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Richardson, Charles
; APPLICANT: Chambers, James
; APPLICANT: Causey, Stuart
; APPLICANT: Pollock, Thomas J.

APPLICANT: Cappello, Joseph
APPLICANT: Crisman, John W.
TITLE OF INVENTION: No. 6018030el Peptides Comprising Repetitive
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,085B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-6/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-398-3249
TELEFAX: 415-781-1989
INFORMATION FOR SEQ ID NO: 91:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-482-085B-91
Query Match 46.6%; Score 61; DB 3; Length 47;
Best Local Similarity 54.5%; Pred. No. 0.22;
Matches 12; Conservative 1; Mismatches 7; Indels 2; Gaps 2;
QY 2 CTGACCTCTTACGCTGCTGTC 21
DB 10 CTCGATCTACATGATGCACC 31
RESULT 15
US-08-900-230-17
; Sequence 17, Application US/08900230
; Patent No. 6329197
; GENERAL INFORMATION:
; APPLICANT: Bard, Jonathan A.
; TITLE OF INVENTION: DNA ENCODING GALANN GALR3 RECEPTORS AND
; TITLE OF INVENTION: USUS THEROP
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of The Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

ZIP: 11036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,230
FILING DATE: 23-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52241-C/JPM/ADM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-900-230-17

Query Match 45.8%; Score 60; DB 4; Length 45;
Best Local Similarity 47.4%; Pred. NO. 0.28;
Matches 9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 3 TGXADCTGCTYACTGCGXC 21
Db 26 TGGGCGATCCACAGAGGC 44

Search completed: October 21, 2002, 16:40:36
Job time : 5.70588 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:36:38 : Search time 4.76471 Seconds
(without alignments)
484.006 Million cell updates/sec

Title: US-10-032-658-4

Perfect score: 131

Sequence: 1 XCTGXADCTSCRYACTGCGXCPNA 24

Scoring table: BLOSUM62

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 69 | 52.7 | 1176 | 2 A49848 | nitrite reductase |
| 2 | 65 | 49.6 | 1104 | 2 JH0181 | nitrite reductase |
| 3 | 60 | 45.8 | 325 | 2 H71271 | hypothetical prote |
| 4 | 59 | 45.0 | 1044 | 2 T43155 | nitrite reductase |
| 5 | 59 | 45.0 | 1044 | 2 T43160 | nitrite reductase |
| 6 | 58 | 44.3 | 719 | 2 T00266 | hypothetical prote |
| 7 | 57 | 43.5 | 63 | 2 S25772 | testis-specific pr |
| 8 | 56.5 | 43.1 | 2135 | 2 T14602 | variant-specific s |
| 9 | 55.5 | 42.4 | 79 | 2 T17014 | metallothionein-11 |
| 10 | 55.5 | 42.4 | 141 | 2 PC1294 | trophozoite surfac |
| 11 | 54 | 41.2 | 807 | 2 E83726 | assimilatory nitri |
| 12 | 53.5 | 40.8 | 230 | 2 A38346 | ultra-high-sulfur |
| 13 | 53 | 40.5 | 25 | 1 SMMR | metallothionein - |
| 14 | 53 | 40.5 | 65 | 2 A38739 | metallothionein - |
| 15 | 53 | 40.5 | 166 | 2 G98188 | ferredoxin-like pr |
| 16 | 53 | 40.5 | 166 | 2 AB3098 | hypothetical prote |
| 17 | 53 | 40.5 | 390 | 2 F69086 | pyruvate formate-1 |
| 18 | 52 | 39.7 | 972 | 2 A30363 | glycoprotein GP330 |
| 19 | 52 | 39.7 | 4660 | 2 T42737 | gp330 protein prec |
| 20 | 51.5 | 39.3 | 72 | 2 T07073 | metallothionein ty |
| 21 | 51.5 | 39.3 | 73 | 2 T07109 | metallothionein-11 |
| 22 | 51.5 | 39.3 | 74 | 2 T16879 | metallothionein-11 |
| 23 | 51.5 | 39.3 | 78 | 2 A23239 | metallothionein-11 |
| 24 | 51.5 | 39.3 | 178 | 2 A23239 | metallothionein-11 |
| 25 | 51.5 | 39.3 | 279 | 2 A81225 | high-cysteine chor |
| 26 | 51.5 | 39.3 | 279 | 2 D81997 | ferredoxin, 4Fe-4S |
| 27 | 51.5 | 39.3 | 787 | 2 T05617 | probable ferredoxi |
| 28 | 51.5 | 39.3 | 1797 | 2 A55677 | hypothetical prote |
| 29 | 51.5 | 39.3 | 1798 | 2 A55677 | laminin beta-2 cna |
| | | | | 2 S53869 | laminin beta-2 cna |

| | | | | | |
|----|------|------|------|----------|--------------------|
| 30 | 51.5 | 39.3 | 2215 | 2 T00348 | LR11 protein - mou |
| 31 | 51 | 38.9 | 60 | 2 B27490 | metallothionein B |
| 32 | 51 | 38.9 | 166 | 2 D95346 | NAPF Ferredoxin co |
| 33 | 51 | 38.9 | 208 | 2 E90926 | probable oxidoredu |
| 34 | 51 | 38.9 | 208 | 2 A85775 | probable oxidoredu |
| 35 | 51 | 38.9 | 208 | 2 B64925 | probable oxidoredu |
| 36 | 51 | 38.9 | 223 | 2 B38346 | ultra-high-sulfur |
| 37 | 51 | 38.9 | 1806 | 2 T23298 | hypothetical prote |
| 38 | 50.5 | 38.5 | 60 | 2 S31723 | metallothionein - |
| 39 | 50.5 | 38.5 | 61 | 2 A37425 | metallothionein 2 |
| 40 | 50.5 | 38.5 | 62 | 2 S54335 | metallothionein-2c |
| 41 | 50.5 | 38.5 | 62 | 2 S54336 | metallothionein-2a |
| 42 | 50.5 | 38.5 | 68 | 2 B46034 | metallothionein 3' |
| 43 | 50.5 | 38.5 | 68 | 2 S44392 | metallothionein 3' |
| 44 | 50.5 | 38.5 | 68 | 2 UC6521 | metallothionein II |
| 45 | 50.5 | 38.5 | 72 | 2 T07105 | metallothionein-11 |

ALIGNMENTS

RESULT 1

A49848

nitrite reductase (NADH) (EC 1.6.6.4) - Neurospora crassa

C:Species: Neurospora crassa

C>Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 12-Sep-1997

C:Accession: A49848

R:Exley, G.E.; Colandene, J.D.; Garrett, R.H.

J. Bacteriol. 175, 2379-2392, 1993

A:Title: Molecular cloning, characterization, and nucleotide sequence of nit-6, the s

A:Reference number: A49848; MUID:93224461

A:Accession: A49848

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1176 <EXL>

A>Note: sequence inconsistent with nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIN:129213, NCBIPI:129214)

C:Keywords: oxidoreductase

Query Match

Best Local Similarity

Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 515 CTSGLDKSCRYACTGCGGC 534

QY 2 CTGXADCTSCRYACTGCGXC 21

JH0181

nitrite reductase (NADH) (EC 1.6.6.4), long form - Emeritella nidulans

C:Species: Emeritella nidulans, Aspergillus nidulans

C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 01-Dec-2000

C:Accession: JH0181; PS0298; PS0299

R:Johnstone, I.L.; McCabe, P.C.; Greaves, P.; Gurr, S.J.; Cole, G.E.; Brow, M.A.D.; U

Gene 90, 181-197, 1990

A:Title: Isolation and characterization of the crna-nia-nad gene cluster for nitrat

A:Reference number: JH0181; MUID:9038264

A:Accession: JH0181

A:Molecule type: DNA

A:Residues: 1-1104 <JH>

A:Accession: PS0298

A:Molecule type: mRNA

A:Residues: 1-1104 <J02>

A:Accession: PS0299

A:Molecule type: mRNA

A:Residues: 1-770, 'E' <J03>

C:Comment: This protein is required for the assimilation of nitrate.

C:Genetics:

A:Gene: niaA

A:introns: 59/2; 109/3; 289/3; 493/2; 561/1; 718/3; 771/2

C:Keywords: alternative splicing; NAD; oxidoreductase

C:Superfamily: fruit fly testis-specific protein
C:Keywords: spermatogenesis; tandem repeat

Query Match 43.5%; Score 57; DB 2; Length 63;
Best Local Similarity 45.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXC 21
| | | | |
Db CGCGCGCGGCGCGCGGCGC 39

RESULT 8
T14602

variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)
C:Species: Plasmodium falciparum

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000

C:Accession: T14602

R:Voss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
submitted to the EMBL Data Library, February 1998

A:Description: Identification of a conserved 5' flanking region of Plasmodium falciparum

A:Reference number: Z18158

A:Accession: T14602

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2135 <VOS>

A:Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AAC05220.1

C:Genetics:

A:Gene: varphi17

Query Match 43.1%; Score 56.5; DB 2; Length 2135;
Best Local Similarity 57.9%; Pred. No. 40;
Matches 11; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 2 CT-GXADCTCTXACTGCGC 19
| | | | |
Db 1133 CTSGDSCTPCACACTTGG 1151

RESULT 9
T17014

metallothionein-like protein AMT1 - apple tree
C:Species: Malus domestica (apple tree)

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000

C:Accession: T17014

R:Reid, S.J.; Ross, G.S.

Physiol. Plantarum 100, 183-189, 1997

A:Title: Up-regulation of two cDNA clones encoding metallothionein-like proteins in apple

A:Reference number: Z18652

A:Accession: T17014

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-79 <REI>

A:Cross-references: EMBL:U61973; NID:g1655850; PID:g1655851

A:Experimental source: apple flesh cortical tissue

C:Genetics:

A:Gene: AMT1

C:Superfamily: metallothionein

C:Keywords: metal binding

Query Match 42.4%; Score 55.5; DB 2; Length 79;
Best Local Similarity 40.0%; Pred. No. 6.3;
Matches 10; Conservative 1; Mismatches 11; Indels 3; Gaps 1;

QY 2 CTGXADC---TSCTXACTGCGXCPN 23
| | | | |
Db 5 CGGCGCGSVSCGSGCGCGMAPD 29

RESULT 10
PC1294
trophozoite surface antigen - Giardia lamblia (strain Adelaide-1) (fragment)
C:Species: Giardia lamblia

C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 09-Sep-1997
C:Accession: PC1294

R:EY, P.L.; Mayrhofer, G.

Gene 129, 257-262, 1993

A:Title: Two genes encoding homologous 70-kDa surface proteins are present within ind

A:Reference number: PC1294; MUID:93314970

A:Molecule type: DNA

A:Accession: PC1294

A:Residues: 1-141 <EYP>

A:Cross-references: GB:M97488; NID:g159129; PID:g159130

C:Genetics:

A:Gene: tsa

C:Keywords: surface antigen

Query Match 42.4%; Score 55.5; DB 2; Length 141;
Best Local Similarity 41.7%; Pred. No. 9.1;
Matches 10; Conservative 1; Mismatches 6; Indels 7; Gaps 1;

QY 2 CTGXADCTSC-----TXACTGC 18
| | | | |
Db 102 CTSSPDCTTCUDGYKASACTKC 125

RESULT 11
E83726

assimilatory nitrite reductase (subunit) nasD [imported] - Bacillus halodurans (strai

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: E83726

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: E83726

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-807 <STO>

A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04332.1; GSPDB:G

C:Genetics:

A:Gene: nasD

Query Match 41.2%; Score 54; DB 2; Length 807;
Best Local Similarity 45.0%; Pred. No. 41;
Matches 9; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXC 21
| | | | |
Db 438 CASVADVGACTAGRSCGTC 457

RESULT 12
A8346

ultra-high-sulfur keratin 1 - mouse

C:Species: Mus musculus (house mouse)

C>Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 24-Sep-1999

C:Accession: A8346

R:Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogeli, G.

J. Biol. Chem. 265, 21375-21380, 1990

A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skl

A:Reference number: A8346; MUID:91065960

A:Accession: A8346

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-230 <WOO>

A:Cross-references: GB:M37759; NID:g200961; PIDN:AAA40106.1; PID:g200962

C:Superfamily: ultra-high-sulfur keratin

Query Match 40.8%; Score 53.5; DB 2; Length 230;
Best Local Similarity 39.1%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 9; Indels 3; Gaps 1;

QY 2 CTGXADCTSC---CTXACTGCGXC 21

Db 55 CVPVSCSSCGCGSCGCGCGC 77

RESULT 13

SMR

metallothionein - cultivated mushroom

C:Species: Agaricus bisporus (cultivated mushroom)

C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 13-Sep-1996

C:Accession: A03286

R:Munger, K.; Lerch, K.

Biochemistry 24, 6751-6756, 1985

A:Title: Copper metallothionein from the fungus Agaricus bisporus: chemical and spectroscopic studies

A:Reference number: A03286

A:Accession: A03286

A:Molecule type: protein

A:Residues: 1-25 <MUN>

A:Experimental source: mycelium; strain A-32

C:Comment: In contrast to the vertebrate metallothioneins, the fungal proteins bind copper

C:Comment: The absorbative, luminescent, and stereoptical properties of the copper MT are

C:Keywords: chelation; metal binding; metal-thiolate cluster

Query Match

Best Local Similarity 40.5%; Score 53; DB 1; Length 25;

Matches 10; Conservative 3; Mismatches 3; Indels 6; Gaps 2;

QY 2 CTGXADCT---SCTXACTGCGC 19

Db 5 CSGASCTCGASGCT--CSGCG 24

RESULT 14

A38739

metallothionein - sea urchin (Strongylocentrotus purpuratus)

C:Species: Strongylocentrotus purpuratus (purple urchin)

C>Date: 13-Sep-1991 #sequence_revision 13-Sep-1991 #text_change 20-Aug-1999

C:Accession: A38739

R:Nemer, M.; Thornton, R.D.; Stuebing, E.W.; Harlow, P.

J. Biol. Chem. 266, 6586-6593, 1991

A:Title: Structure, spatial, and temporal expression of two sea urchin metallothionein

A:Reference number: A38739; PMID:91177920

A:Accession: A38739

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-65 <NEM>

A:Cross-references: GB:M59822; NID:g161535; PIDN:AAA30062.1; PID:g161536

C:Superfamily: metallothionein

Query Match

Best Local Similarity 40.5%; Score 53; DB 2; Length 65;

Matches 12; Conservative 0; Mismatches 11; Indels 2; Gaps 1;

QY 2 CTGXADCT--SCTXACTGCGCPNA 24

Db 17 CTGDDCTTGCKCKGCGCSNA 41

RESULT 15

G98188

ferrodoxin-like protein (AF083948) [imported] - Agrobacterium tumefaciens (strain C58, C)

C:Species: Agrobacterium tumefaciens

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002

C:Accession: G98188

R:Godner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: G98188

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-166 <KUR>

A:Cross-references: GB:AB07870; PIDN:AAK89033.1; PID:g15158827; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_920
A:Map position: linear chromosome

Query Match

Best Local Similarity 40.5%; Score 53; DB 2; Length 166;

Matches 11; Conservative 1; Mismatches 6; Indels 24; Gaps 1;

QY 4 GXADCTSC-----TXACTGCGXC 21

Db 104 GAVDCOACRDACPTAIRPRRGSPVPAIVEDACTCGAC 145

Search completed: October 21, 2002, 16:40:09
Job time : 6.76471 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:23 : Search time 2.82353 Seconds
(without alignments)
329.116 million cell updates/sec

Title: US-10-032-658-4

Perfect score: 131

Sequence: 1 XCTGXADCTSCFXACTGCXCPNA 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 69 | 52.7 | 1176 | 1 | NIR_NEUCR |
| 2 | 65 | 49.6 | 1104 | 1 | NIR_EMENT |
| 3 | 60 | 45.8 | 325 | 1 | Y856_TREPA |
| 4 | 57 | 43.5 | 63 | 1 | M84A_DROME |
| 5 | 55.5 | 42.4 | 79 | 1 | MT2_MALDO |
| 6 | 53 | 40.5 | 25 | 1 | MT_AGABT |
| 7 | 53 | 40.5 | 65 | 1 | MTB_STRPU |
| 8 | 52 | 39.7 | 60 | 1 | MTB_CHIAC |
| 9 | 52 | 39.7 | 60 | 1 | MTB_CHIHA |
| 10 | 52 | 39.7 | 60 | 1 | MTB_PAGBE |
| 11 | 52 | 39.7 | 60 | 1 | MTB_LIZAV |
| 12 | 52 | 39.7 | 60 | 1 | LRP2_RAT |
| 13 | 51.5 | 39.3 | 60 | 1 | MTA_SPAAU |
| 14 | 51.5 | 39.3 | 72 | 1 | MT2A_LYCES |
| 15 | 51.5 | 39.3 | 73 | 1 | MT2Y_LYCES |
| 16 | 51.5 | 39.3 | 74 | 1 | MT2_NICPL |
| 17 | 51.5 | 39.3 | 78 | 1 | MT2_ACRCH |
| 18 | 51.5 | 39.3 | 178 | 1 | CHHC_BOOMO |
| 19 | 51.5 | 39.3 | 194 | 1 | KRUB_HUMAN |
| 20 | 51.5 | 39.3 | 1798 | 1 | LMB2_HUMAN |
| 21 | 51.5 | 39.3 | 2215 | 1 | MTB_ONCMY |
| 22 | 51 | 38.9 | 60 | 1 | SORL_MOUSE |
| 23 | 51 | 38.9 | 1134 | 1 | VGLM_HANTH |
| 24 | 50.5 | 38.5 | 60 | 1 | MT_AMBME |
| 25 | 50.5 | 38.5 | 60 | 1 | MT_ESOUL |
| 26 | 50.5 | 38.5 | 62 | 1 | MT2A_RABIT |
| 27 | 50.5 | 38.5 | 62 | 1 | MT2C_RABIT |
| 28 | 50.5 | 38.5 | 68 | 1 | MT3_HORSE |
| 29 | 50.5 | 38.5 | 68 | 1 | MT3_HUMAN |
| 30 | 50.5 | 38.5 | 68 | 1 | MT3_PIG |
| 31 | 50.5 | 38.5 | 72 | 1 | MT2X_LYCES |
| 32 | 50.5 | 38.5 | 713 | 1 | TS44_GIALA |
| 33 | 50.5 | 38.5 | 1291 | 1 | YC81_CABEL |

| | | | | | | |
|----|------|------|------|---|------------|--------------------|
| 34 | 50.5 | 38.5 | 1592 | 1 | SORL_CHICK | Q98930 g sortilin- |
| 35 | 50.5 | 38.5 | 1799 | 1 | LMB2_MOUSE | O61292 mus musculu |
| 36 | 50.5 | 38.5 | 1801 | 1 | LMB2_RAT | P15800 rattus norv |
| 37 | 50 | 38.2 | 60 | 1 | MTB_DICLA | O9PL99 dicentrarch |
| 38 | 50 | 38.2 | 60 | 1 | MTB_SALSA | P52720 salmo salar |
| 39 | 50 | 38.2 | 60 | 1 | MT_PERFL | P52725 perca fluvi |
| 40 | 50 | 38.2 | 60 | 1 | MT_ZOAVI | P52728 zoares viv |
| 41 | 50 | 38.2 | 74 | 1 | M84B_DROME | O01643 drosophila |
| 42 | 50 | 38.2 | 279 | 1 | NAPG_HAEIN | P44652 haemophilus |
| 43 | 49.5 | 37.8 | 34 | 1 | SCAM_SCOMA | P80719 scorpio mau |
| 44 | 49.5 | 37.8 | 60 | 1 | MT2_CYPCA | O91910 cyprinus ca |
| 45 | 49.5 | 37.8 | 60 | 1 | MT_BRAKE | P52722 brachydanio |

ALIGNMENTS

| RESULT 1 | ID | NIR_NEUCR | STANDARD: | PRT: 1176 AA. |
|----------|--|-----------|-----------|---------------|
| AC | P38681: | | | |
| DT | 01-FEB-1995 (Rel. 31, Created) | | | |
| DT | 01-FEB-1995 (Rel. 31, Last sequence update) | | | |
| DT | 15-JUL-1999 (Rel. 38, Last annotation update) | | | |
| DE | Nitrite reductase [NAD(P)H] (EC 1.6.6.4). | | | |
| GN | NIT-6. | | | |
| OS | Neurospora crassa. | | | |
| OC | Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; | | | |
| OX | Sordariales; Sordariaceae; Neurospora. | | | |
| RN | NCBI_TaxID=5141; | | | |
| RP | (1) | | | |
| RC | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-74-OR2-1A: | | | |
| RA | MEDLINE-93224461; PubMed-8096840; | | | |
| RA | Exley G.E., Colandene J.D., Garrett R.H.: | | | |
| RT | "Molecular cloning, characterization, and nucleotide sequence of | | | |
| RT | nit-6, the structural gene for nitrite reductase in Neurospora | | | |
| RT | crassa."; | | | |
| RT | J. Bacteriol. 175:2379-2392(1993). | | | |
| CC | -1- CATALYTIC ACTIVITY: 3 NAD(P)H + nitrite -> 3 NAD(P)(+) + NH(4)OH + H(2)O. | | | |
| CC | -1- COFACTOR: THIS ENZYME IS A FAD FLAVOPROTEIN THAT ALSO CONTAINS | | | |
| CC | A SIROHEME AND ONE 2FE-2S IRON-SULFUR CENTER. | | | |
| CC | -1- PATHWAY: SECOND STEP IN NITRATE ASSIMILATION (DENITRIFICATION). | | | |
| CC | -1- SUBUNIT: HOMODIMER (BY SIMILARITY). | | | |
| CC | -1- INDUCTION: BY NITRATE. | | | |
| CC | -1- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN | | | |
| CC | FOUND IN NITRATE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND | | | |
| CC | SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1). | | | |
| CC | ----- | | | |
| CC | THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION | | | |
| CC | BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - | | | |
| CC | THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS | | | |
| CC | USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY | | | |
| CC | MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL | | | |
| CC | ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/ | | | |
| CC | OR SEND AN EMAIL TO license@sib-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL: L07391; -; NOT_ANNOTATED_CDS. | | | |
| DR | PIR: A49848; A49848. | | | |
| DR | InterPro: IPR001327; FAD_Pyr_redox. | | | |
| DR | InterPro: IPR000660; Nit_Sir. | | | |
| DR | InterPro: IPR001281; Rieske. | | | |
| DR | Pfam: PR01070; NIT_SIR; 1. | | | |
| DR | Pfam: PR00070; Pyr_redox; 1. | | | |
| DR | Pfam: PF00355; Rieske; 1. | | | |
| DR | PRINTS: PR00397; SIROHAEM. | | | |
| DR | PROSITE: PS00365; NIT_SIR; 1. | | | |
| KW | Oxidoreductase; FAD; Flavoprotein; Iron-sulfur; Nitrate assimilation; | | | |
| KW | Heme; NADP. | | | |
| FT | NP_BIND 26 60 FAD (POTENTIAL); | | | |
| FT | NP_BIND 183 215 NAD(P)H (POTENTIAL); | | | |
| FT | METAL 717 717 IRON-SULFUR (2FE-2S) (POTENTIAL); | | | |

```

FT METAL 723 723 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 757 757 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 761 761 IRON-SULFUR (2FE-2S) AND SIROHEME
FT DOMAIN 998 1054 (BY SIMILARITY).
SQ SEQUENCE 1176 AA; 127367 MW; FFC7DCE66F80C710 CRC64;

Query Match 52.7%; Score 69; DB 1; Length 1176;
Best Local Similarity 60.0%; Pred. No. 0.072;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 CTGAXDCTCTXACTGCGXC 21
DB 515 CTSLGDLKCTKAGTCGCGC 534

RESULT 2
NIR_EMBL1 STANDARD; PRT; 1104 AA.
AC P22944;
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Nitrite reductase [NAD(P)H] (EC 1.6.6.4).
GN NITR.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Emericella.
ON NCBI_TaxID=5072;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=90382664; PubMed=2205530;
RA Johnson E.L., McCabe P.C., Greaves P., Gurr S.J., Cole G.E.,
RA Brown A.D., Umiles S.E., Clutterbuck A.J., Kinghorn J.R., Innis M.A.;
RA "Isolation and characterisation of the crnA-nla-nlaD gene cluster
RT for nitrate assimilation in Aspergillus nidulans.";
RL Gene 90:181-192(1990).
CC -1- CATALYTIC ACTIVITY: 3 NAD(P)H + nitrite = 3 NAD(P)(+) + NH(4)OH +
CC H(2)O.
CC -1- COFACTOR: THIS ENZYME IS A FAD FLAVOPROTEIN THAT ALSO CONTAINS
CC A SIROHEME AND ONE 2FE-2S IRON-SULFUR CENTER.
CC -1- PATHWAY: SECOND STEP IN NITRATE ASSIMILATION (DENITRIFICATION).
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN
CC FOUND IN NITRATE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
CC SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M58289; AAA33315.1; -.
DR PIR; JH0181.
DR InterPro; IPR001327; FAD_DYF_redox.
DR InterPro; IPR000660; NIT_Sir.
DR Pfam; PF01077; NIR_SIR; 1.
DR Pfam; PF00070; PYR_redox; 1.
DR Pfam; PF00355; Rlseske; 1.
DR PRINTS; PR00397; SIROHEM.
DR PROSITE; PS00365; NIR_SIR; 1.
KW Oxidoreductase; FAD; Flavoprotein; Iron-sulfur; Nitrate assimilation;
KW Heme; NADP.
FT NP_BIND 44 79 FAD (POTENTIAL).
FT NP_BIND 146 176 NAD(P)H (POTENTIAL).
FT METAL 720 720 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 726 726 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 760 760 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 764 764 IRON-SULFUR (2FE-2S) AND SIROHEME

```

```

FT SQ SEQUENCE 1104 AA; 122731 MW; E5D00436E1E86304 CRC64;
SQ SEQUENCE 1104 AA; 122731 MW; E5D00436E1E86304 CRC64;

Query Match 49.6%; Score 65; DB 1; Length 1104;
Best Local Similarity 60.0%; Pred. No. 0.22;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 CTGAXDCTCTXACTGCGXC 21
DB 519 CKTIADVKCTKAGTCGCGC 538

RESULT 3
Y856_TREPA STANDARD; PRT; 325 AA.
AC O83828;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein TP0856 precursor.
GN TP0856.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
ON NCBI_TaxID=160;
RN [1]
RX SEQUENCE FROM N.A.
RA STRAIN=NICHOLS.
RA MEDLINE=9832770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -1- SIMILARITY: BELONGS TO THE UPF0164 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001256; AAC65828.1; -.
DR TIGR; TP0856; -.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 325 HYPOTHETICAL PROTEIN TP0856.
SQ SEQUENCE 325 AA; 34054 MW; F9CFDCBD253C07D2 CRC64;

Query Match 45.8%; Score 60; DB 1; Length 325;
Best Local Similarity 47.6%; Pred. No. 0.38;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 3 TGXADTCTXACTGCGXCPN 23
DB 233 TGEQCKCTKSGRCGTGTCNCN 253

RESULT 4
M84A_DROME STANDARD; PRT; 63 AA.
AC 001642; Q9VIA3;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Male specific sperm protein Mst84A.
GN Mst84A OR CGI7946.

```

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=92102953; PubMed=1684716;
 RA Kuhn R., Kuhn C., Boersch D., Glaetzer K.H., Schaefer U.,
 RT "A cluster of four genes selectively expressed in the male germ line
 of Drosophila melanogaster.";
 RL Mech. Dev. 35:143-151(1991).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson R., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Howack J.,
 RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Ibegwam C.,
 RA Jatalil B.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacלב J.M.,
 RA Palazkzo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA She B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2183-2195(2000).
 CC -1- TISSUE SPECIFICITY: TESTIS.
 CC -1- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.
 CC -1- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P
 MOTIFS.
 CC -1- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X67703; CAA47937.1; -
 DR EMBL: AE003672; AAF54022.1; -
 DR HSSP: P01180; INPO.

DR FlyBase: FBgn0004172; Mst84Da.
 KW Spermatogenesis; Repeat; Multigene family.
 SQ SEQUENCE 63 AA; 5806 MW; BF84CD74CBCEFD3F CRC64;
 Query Match 43.5%; Score 57; DB 1; Length 63;
 Best Local Similarity 45.0%; Pred. No. 0.27;
 Matches 9; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 2 CTGXADCTCTXACTGCGXC 21
 Db 20 CGGCGCGCGCGCGCGCGC 39
 RESULT 5
 WT2_MALDO STANDARD; PRT; 79 AA.
 ID WT2_MALDO
 AC 024058;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Metallothionein-like protein type 2.
 GN MT1.
 OS Malus domestica (apple) (Malus sylvestris).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Rosales; Rosaceae; Maloideae; Malus.
 NC NCB1_TaxID=3750;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fruit cortical tissue;
 RA Reid S.J., Ross G.S.;
 RT "Up-regulation of two cDNA clones encoding metallothionein-like
 RT proteins in apple fruit during cool storage.";
 RL Physiol. Plantarum 100:183-189(1997).
 CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS.
 CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U61973; AAC23697.1; -
 DR InterPro: IPR000347; Metallothion_15.
 DR Pfam: PF01439; Metallothio_2; 1.
 DR ProDom: PD001611; Metallothion.15; 1.
 KW Metal-binding; Metal-thiolate cluster; Chelation.
 SQ SEQUENCE 79 AA; 7836 MW; 8ADC58BD8B644CC CRC64;
 Query Match 42.4%; Score 55.5; DB 1; Length 79;
 Best Local Similarity 40.0%; Pred. No. 0.49;
 Matches 10; Conservative 1; Mismatches 11; Indels 3; Gaps 1;
 QY 2 CTGXADC--TSCTXACTGCGXCPN 23
 Db 5 CGGKCGGCGVSCGSGCGCGMAPD 29
 RESULT 6
 MT_AGABI
 ID MT_AGABI STANDARD; PRT; 25 AA.
 AC P04358;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Metallothionein (MT).
 OS Agaricus bisporus (Common mushroom).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Agaricaceae; Agaricus.

```

OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE.
RA STRAIN=A-32; TISSUE=Mycellium;
RC Muenger K., Lerch K.;
RT "Copper metallothionein from the fungus Agaricus bisporus: chemical
RL and spectroscopic properties.";
RL Biochemistry 24:6751-6756(1985).
CC -1- FUNCTION: THE METALLOTHIONEINS ARE INVOLVED IN THE CELLULAR
CC SEQUESTRATION OF TOXIC METAL IONS. BINDS SIX COPPER (COPROUS)
CC IONS.
CC -1- MISCELLANEOUS: THE ABSORPTION, LUMINESCENT, AND STEREOPTICAL
CC PROPERTIES OF THE COPPER MT ARE ATTRIBUTED TO THE METAL-THIOLATE
CC COMPLEX BECAUSE THEY ARE NOT PRESENT IN THE APOPROTEIN.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 8.
DR PIR: A03286; SMMR.
RW Metal-binding; Metal-thiolate cluster; Chelation; Copper.
SQ SEQUENCE 25 AA; 2233 MW; 33AB70F21023CFF7 CRC64;

OY 2 CTGXADCT---SCTXACTGCG 19
Db 1:1 11 11 1:111
5 CSGASSCTGCASGCT--CSGCC 24

RESULT 7
MTB_STRPU STANDARD: PRT: 65 AA.
AC 027287;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Metallothionein-B (MTB).
GN MTB1 OR MTB.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Echinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OX Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87172743; PubMed=3561398;
RA Wilkinson D.G., Nemer M.;
RT "Metallothionein genes MTA and MTB expressed under distinct
RT quantitative and tissue-specific regulation in sea urchin embryos.";
RL Mol. Cell. Biol. 7:48-58(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=91177920; PubMed=2007604;
RA Nemer M., Thornton R.D., Stuebing E.W., Harlow P.;
RT "Structure, spatial, and temporal expression of two sea urchin
RT metallothionein genes, SPMTB1 and SPMTA.";
RL J. Biol. Chem. 266:6586-6593(1991).
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC -1- INDUCTION: BY HEAVY METALS.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: M15989; AAA30063.1; -
DR EMBL: M59822; AAA30062.1; -.
DR HSSP: P04734; LOUL.
DR InterPro: IPR001396; Metallothion_Echnd.

```

```

DR PRINTS:PR00873; MTECHINOIDEA.
KW Metal-binding; Metal-thiolate cluster; Chelation.
SQ SEQUENCE 65 AA; 6470 MW; 12EIC08064D9121A5 CRC64;

Query Match 40.5%; Score 53; DB 1; Length 65;
Best Local Similarity 48.0%; Pred. No. 0.87;
Matches 12; Conservative 0; Mismatches 11; Indels 2; Gaps 1;

Oy 2 CTGXADCT--SCTXACTGCGXCPNA 24
||| ||| | ||| |||
Db 17 CTGDDCTTGKCKCKDCTCGKCSNA 41

RESULT 8
MTB_CHAAC
AC ID MTB_CHAAC STANDARD: PRT: 60 AA.
P52724;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Metallothionein B (MT-B).
GN MTB.
OS Chenocephalus aceratus (White crocodile fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorphi; Perciformes;
OC Noronhenioidae; Channichthyidae; Chenocephalus.
OX NCBI_TaxId=36190;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RC MEDLINE=97233027; PubMed=9078263;
RA Scudiero R., Carignale V., Riggio M., Capasso C., Capasso A.,
RA Kille P., di Prisco G., Parisi E.;
RT "Difference in hepatic metallothionein content in Antarctic
RT red-blooded and haemoglobinless fish: undetectable metallothionein
RT levels in haemoglobinless fish is accompanied by accumulation of
RT untranslated metallothionein mRNA.";
RL Biochem. J. 322:207-211(1997).

-1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
-1- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
-1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: 272483; CA96564.1; -.
DR HSSP: P02802; IDRS.
DR InterPro: IPR003019; Metallthion.
DR pfam: PF00131; metallo; 1.
DR PRINTS: PR00860; MTEVERTBRATE.
DR PROSITE: PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster; Chelation.
FT DOMAIN 1 28 BETA.
FT METAL 29 60 ALPHA.
FT METAL 4 4 CLUSTER B.
FT METAL 6 6 CLUSTER B.
FT METAL 12 12 CLUSTER B.
FT METAL 14 14 CLUSTER B.
FT METAL 18 18 CLUSTER B.
FT METAL 20 20 CLUSTER B.
FT METAL 23 23 CLUSTER B.

```



```

FT METAL 25 25 CLUSTER B.
FT METAL 28 28 CLUSTER B.
FT METAL 32 32 CLUSTER A.
FT METAL 33 33 CLUSTER A.
FT METAL 35 35 CLUSTER A.
FT METAL 36 36 CLUSTER A.
FT METAL 40 40 CLUSTER A.
FT METAL 43 43 CLUSTER A.
FT METAL 47 47 CLUSTER A.
FT METAL 49 49 CLUSTER A.
FT METAL 54 54 CLUSTER A.
FT METAL 58 58 CLUSTER A.
FT METAL 59 59 CLUSTER A.
SQ SEQUENCE 60 AA; 6006 MW; EF61F4AD61B9124A CRC64;

Query Match 39.7%; Score 52; DB 1; Length 60;
Best Local Similarity 43.5%; Pred. No. 1.1;
Matches 10; Conservative 2; Mismatches 7; Indels 4; Gaps 2;

OY 2 CTGXADCTSCYXACTGC--GXCP 22
Db 14 CGGSCCTCNC--SCTSKKSCCP 34

RESULT 9
MTB_CHIHA STANDARD; PRT; 60 AA.
AC 013259;
AT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Metallothionein B (MT-B).
GN MTB.
OS Chionodraco hamatus (Antarctic teleost icefish),
OS Morone saxatilis (Striped bass), and
OS Gymnodraco acuticeps (Antarctic dragonfish).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidae; Channichthyidae; Chionodraco.
OX NCBI_TaxID=36188, 34816, 8218;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=C.hamatus; TISSUE=Liver;
RA MEDLINE=98264843; PubMed=9601077;
RA Carginalo V., Scudiero R., Capasso C., Kille P.,
RA di Prisco G., Parisi E.;
RT "Cadmium-induced differential accumulation of metallothionein isoforms
RT in the Antarctic icefish, which exhibits no basal metallothionein
RT protein but high endogenous mRNA levels.";
RL Blochem. J. 332:475-481(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=G.acuticeps; TISSUE=Liver;
RA MEDLINE=99334610; PubMed=10406107;
RA Bargelloni L., Scudiero R., Parisi E., Carginalo V., Capasso C.,
RA Patarello T.;
RT "Metallothioneins in antarctic fish: evidence for independent
RT duplication and gene conversion.";
RL Mol. Biol. Evol. 16:885-897(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [12]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [13]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [14]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [15]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [16]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [17]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [18]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [19]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [20]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [21]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [22]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [23]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [24]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [25]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [26]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [27]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [28]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [29]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [30]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [31]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [32]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [33]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [34]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [35]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [36]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [37]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [38]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [39]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [40]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [41]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [42]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [43]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [44]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [45]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [46]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [47]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [48]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [49]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [50]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [51]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [52]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [53]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [54]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [55]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [56]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [57]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [58]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [59]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [60]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [61]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [62]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [63]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [64]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [65]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [66]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [67]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [68]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [69]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [70]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [71]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [72]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [73]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [74]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [75]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [76]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [77]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [78]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [79]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [80]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [81]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [82]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [83]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [84]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [85]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [86]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [87]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [88]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [89]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [90]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [91]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [92]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [93]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [94]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [95]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [96]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [97]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [98]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [99]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [100]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leduc G.M., Leduc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.

```

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC
CC EMBL: Y12581; CAA73160.1;
CC EMBL: AF091100; AAC62501.1;
CC EMBL: AJ007561; CAA07556.1;
CC HSSP: P02802; IDES.
CC InterPro: IPR003019; Metallothion.
CC InterPro: IPR000006; Metallothion_vert.
CC Pfam: PF00131; metalthio.1.
CC PRINTS: PR00860; MVERTERRATE.
CC PROSITE: PS00203; METALLOTHIONEIN_VRT.1.
CC Metal-binding; Metal-chiolate cluster; Chelation.
CC
CC METAL 29 60 ALPHA.
CC METAL 4 4 CLUSTER B (BY SIMILARITY).
CC METAL 6 6 CLUSTER B (BY SIMILARITY).
CC METAL 12 12 CLUSTER B (BY SIMILARITY).
CC METAL 14 14 CLUSTER B (BY SIMILARITY).
CC METAL 18 18 CLUSTER B (BY SIMILARITY).
CC METAL 20 20 CLUSTER B (BY SIMILARITY).
CC METAL 23 23 CLUSTER B (BY SIMILARITY).
CC METAL 25 25 CLUSTER B (BY SIMILARITY).
CC METAL 28 28 CLUSTER B (BY SIMILARITY).
CC METAL 32 32 CLUSTER A (BY SIMILARITY).
CC METAL 33 33 CLUSTER A (BY SIMILARITY).
CC METAL 35 35 CLUSTER A (BY SIMILARITY).
CC METAL 36 36 CLUSTER A (BY SIMILARITY).
CC METAL 40 40 CLUSTER A (BY SIMILARITY).
CC METAL 43 43 CLUSTER A (BY SIMILARITY).
CC METAL 47 47 CLUSTER A (BY SIMILARITY).
CC METAL 49 49 CLUSTER A (BY SIMILARITY).
CC METAL 54 54 CLUSTER A (BY SIMILARITY).
CC METAL 58 58 CLUSTER A (BY SIMILARITY).
CC METAL 59 59 CLUSTER A (BY SIMILARITY).
CC SEQUENCE 60 AA; 5978 MW; 5966F5AC61BC424B CRC64;

Query Match 39.7%; Score 52; DB 1; Length 60;
Best Local Similarity 43.5%; Pred. No. 1.1;
Matches 10; Conservative 2; Mismatches 7; Indels 4; Gaps 2;

OY 2 CTGXADCTSCYXACTGC--GXCP 22
Db 14 CGGSCCTCNC--SCTSKKSCCP 34

RESULT 10
MTB_PAGBE STANDARD; PRT; 60 AA.
ID 092145;
AT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Metallothionein B (MT-B).
GN MTB.
OS pagochenia bernacchii (Emerald rockcod) (Trematomus bernacchii),
OS Chionodraco rasbrosphosus (Ocellated icefish),
OS Notothenia coriiceps neglecta (Black rockcod) (Trematomus borchgrevinkii),
OS pagochenia borchgrevinkii (Bald rockcod) (Trematomus borchgrevinkii),
OS and Paracheanichthys charcoti.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidae; Nototheniidae; Trematomus.
OX NCBI_TaxID=40690, 34790, 8209, 8213, 36187;
RN [1]
RP SEQUENCE FROM N.A.

```

RC SPECIES-P.bernacchi, and C.rastrospinosus; TISSUE-Liver;
 RA MEDLINE=97233027; PubMed=9078263;
 RX Scudiero R., Carginale V., Riggio M., Capasso C., Capasso A.,
 RA Kille P., di Prisco G., Parisi E.,
 RT "Difference in hepatic metallothionein content in Antarctic
 red-blooded and haemoglobinless fish: undetectable metallothionein
 RT levels in haemoglobinless fish is accompanied by accumulation of
 RT untranslated metallothionein mRNA.";
 RL Blochem. J. 322:207-211(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=N.c.neglecta, P.borchgrevinkii, and P.charcotii;
 RC TISSUE=Liver;
 RX MEDLINE=99334610; PubMed=10406107;
 RA Bargelloni L., Scudiero R., Parisi E., Carginale V., Capasso C.,
 RA Patarnello T.;
 RT "Metallothioneins in antarctic fish: evidence for independent
 RT duplication and gene conversion";
 RL Mol. Biol. Evol. 16:885-897(1999).
 CC -I- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
 CC -I- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -I- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY, FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z72485; CA96566.1; -;
 DR EMBL; Z72484; CA96565.1; -;
 DR EMBL; AJ006485; CA907064.1; -;
 DR EMBL; AJ007563; CA907558.1; -;
 DR EMBL; AJ007951; CA907786.1; -;
 DR HSSP; P02802; IDPS.
 DR InterPro; IPR003019; Metallthion.
 DR InterPro; IPR000006; Metallthion_vert.
 DR Pfam; PF00131; metalchio; 1.
 DR PRINTS; PR00860; MTVERTEBRATE.
 DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
 KW Metal-binding; Metal-thiolate cluster; Chelation.
 FT DOMAIN 1 28
 FT METAL 29 60 ALPHA.
 FT METAL 4 4 CLUSTER B.
 FT METAL 6 6 CLUSTER B.
 FT METAL 12 12 CLUSTER B.
 FT METAL 14 14 CLUSTER B.
 FT METAL 18 18 CLUSTER B.
 FT METAL 20 20 CLUSTER B.
 FT METAL 23 23 CLUSTER B.
 FT METAL 25 25 CLUSTER B.
 FT METAL 28 28 CLUSTER B.
 FT METAL 32 32 CLUSTER A.
 FT METAL 33 33 CLUSTER A.
 FT METAL 35 35 CLUSTER A.
 FT METAL 36 36 CLUSTER A.
 FT METAL 40 40 CLUSTER A.
 FT METAL 43 43 CLUSTER A.
 FT METAL 47 47 CLUSTER A.
 FT METAL 49 49 CLUSTER A.
 FT METAL 54 54 CLUSTER A.
 FT METAL 58 58 CLUSTER A.
 FT METAL 59 59 CLUSTER A.
 SQ SEQUENCE 60 AA; 5992 MW; E866F4AD61BC424A CRC64;

Best Local Similarity 43.5%; Pred. No. 1.1;
 Matches 10; Conservative 2; Mismatches 7; Indels 4; Gaps 2;
 QY 2 CFCXADCTSTACTGCG--GXCP 22
 DB 14 CGGSGCTCTGNC--SCTSCKNSCCP 34
 RESULT 11
 MT_LIZAU
 ID MT_LIZAU STANDARD; PRT; 60 AA.
 AC O13257;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Metallothionein (MT).
 GN MT.
 OS Liza aurata (Golden grey mullet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Mugilomorpha; Mugilidae;
 OC Liza
 OX NCBI_TaxID=48191;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Cousinou M., Lopez-Barrea J., Dorado G.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBP databases.
 CC -I- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
 CC -I- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -I- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY, FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U93207; AAB51591.1; -;
 DR HSSP; P02802; IDPS.
 DR InterPro; IPR003019; Metallthion.
 DR InterPro; IPR000006; Metallthion_vert.
 DR Pfam; PF00131; metalchio; 1.
 DR PRINTS; PR00860; MTVERTEBRATE.
 DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
 KW Metal-binding; Metal-thiolate cluster; Chelation.
 FT DOMAIN 1 28
 FT METAL 29 60 ALPHA.
 FT METAL 4 4 CLUSTER B.
 FT METAL 6 6 CLUSTER B.
 FT METAL 12 12 CLUSTER B.
 FT METAL 14 14 CLUSTER B.
 FT METAL 18 18 CLUSTER B.
 FT METAL 20 20 CLUSTER B.
 FT METAL 23 23 CLUSTER B.
 FT METAL 25 25 CLUSTER B.
 FT METAL 28 28 CLUSTER B.
 FT METAL 32 32 CLUSTER A.
 FT METAL 33 33 CLUSTER A.
 FT METAL 35 35 CLUSTER A.
 FT METAL 36 36 CLUSTER A.
 FT METAL 40 40 CLUSTER A.
 FT METAL 43 43 CLUSTER A.
 FT METAL 47 47 CLUSTER A.
 FT METAL 49 49 CLUSTER A.
 FT METAL 54 54 CLUSTER A.


```

FT DOMAIN 1932 1972 LDL-RECEPTOR CLASS B 18.
FT DOMAIN 1973 2013 LDL-RECEPTOR CLASS B 19.
FT DOMAIN 2019 2060 EGF-LIKE 8.
FT DOMAIN 2108 2156 LDL-RECEPTOR CLASS B 20.
FT DOMAIN 2158 2201 LDL-RECEPTOR CLASS B 21.
FT DOMAIN 2203 2245 LDL-RECEPTOR CLASS B 22.
FT DOMAIN 2247 2289 LDL-RECEPTOR CLASS B 23.
FT DOMAIN 2291 2332 LDL-RECEPTOR CLASS B 24.
FT DOMAIN 2343 2384 EGF-LIKE 9.
FT DOMAIN 2432 2477 LDL-RECEPTOR CLASS B 25.
FT DOMAIN 2479 2518 LDL-RECEPTOR CLASS B 26.
FT DOMAIN 2520 2562 LDL-RECEPTOR CLASS B 27.
FT DOMAIN 2564 2604 LDL-RECEPTOR CLASS B 28.
FT DOMAIN 2605 2647 LDL-RECEPTOR CLASS B 29.
FT DOMAIN 2652 2694 EGF-LIKE 10.
FT DOMAIN 2699 2739 LDL-RECEPTOR CLASS A 16.
FT DOMAIN 2740 2778 LDL-RECEPTOR CLASS A 17.
FT DOMAIN 2779 2820 LDL-RECEPTOR CLASS A 18.
FT DOMAIN 2821 2862 LDL-RECEPTOR CLASS A 19.
FT DOMAIN 2863 2903 LDL-RECEPTOR CLASS A 20.
FT DOMAIN 2906 2947 LDL-RECEPTOR CLASS A 21.
FT DOMAIN 2948 2992 LDL-RECEPTOR CLASS A 22.
FT DOMAIN 2993 3031 LDL-RECEPTOR CLASS A 23.
FT DOMAIN 3032 3072 LDL-RECEPTOR CLASS A 24.
FT DOMAIN 3073 3112 LDL-RECEPTOR CLASS A 25.
FT DOMAIN 3113 EGF-LIKE 11.
FT DOMAIN 3154 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3241 3282 LDL-RECEPTOR CLASS B 30.
FT DOMAIN 3284 3333 LDL-RECEPTOR CLASS B 31.
FT DOMAIN 3335 3377 LDL-RECEPTOR CLASS B 32.
FT DOMAIN 3379 3420 LDL-RECEPTOR CLASS B 33.
FT DOMAIN 3421 3461 LDL-RECEPTOR CLASS B 34.
FT DOMAIN 3467 3511 EGF-LIKE 13.
FT DOMAIN 3512 3552 LDL-RECEPTOR CLASS A 26.
FT DOMAIN 3553 3593 LDL-RECEPTOR CLASS A 27.
FT DOMAIN 3594 3634 LDL-RECEPTOR CLASS A 28.
FT DOMAIN 3635 3675 LDL-RECEPTOR CLASS A 29.
FT DOMAIN 3678 3718 LDL-RECEPTOR CLASS A 30.
FT DOMAIN 3719 3758 LDL-RECEPTOR CLASS A 31.
FT DOMAIN 3759 3797 LDL-RECEPTOR CLASS A 32.
FT DOMAIN 3798 3836 LDL-RECEPTOR CLASS A 33.
FT DOMAIN 3842 3882 LDL-RECEPTOR CLASS A 34.
FT DOMAIN 3883 3924 LDL-RECEPTOR CLASS A 35.
FT DOMAIN 3928 3966 LDL-RECEPTOR CLASS A 36.
FT DOMAIN 3968 4008 EGF-LIKE 14.
FT DOMAIN 4009 4050 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4156 4197 LDL-RECEPTOR CLASS B 35.
FT DOMAIN 4199 4241 LDL-RECEPTOR CLASS B 36.
FT DOMAIN 4244 4284 LDL-RECEPTOR CLASS B 37.
FT DOMAIN 4332 4370 EGF-LIKE 16.
FT DOMAIN 4379 4413 EGF-LIKE 17.
FT SITE 4457 4463 SH3-BINDING (POTENTIAL).
FT SITE 4454 4460 SH3-BINDING (POTENTIAL).
FT SITE 4457 4463 SH2-BINDING (POTENTIAL).
FT SITE 4619 4625 SH3-BINDING (POTENTIAL).
FT SITE 4624 4630 SH3-BINDING (POTENTIAL).
FT SITE 1743 1745 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 4522 4527 ENDOCYTOSIS SIGNAL (POTENTIAL).
FT SITE 4601 4606 ENDOCYTOSIS SIGNAL (POTENTIAL).
FT DISULFID 28 40 BY SIMILARITY.
FT DISULFID 35 53 BY SIMILARITY.
FT DISULFID 47 62 BY SIMILARITY.
FT DISULFID 67 80 BY SIMILARITY.
FT DISULFID 74 93 BY SIMILARITY.
FT DISULFID 87 103 BY SIMILARITY.
FT DISULFID 108 120 BY SIMILARITY.
FT DISULFID 115 133 BY SIMILARITY.
FT DISULFID 127 142 BY SIMILARITY.
FT DISULFID 147 157 BY SIMILARITY.

```

Query Match 39.7%; Score 52; DB 1; Length 4660;
 Best Local Similarity 36.4%; Pred. No. 27;
 Matches 8; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

```

OY 2 CIGXADCTSCYACTGCGXCPN 23
Db 1320 CCGYSTCINLSALCDGVFDCPN 1341

RESULT 13
MTA_SPAU
ID MTA_SPAU STANDARD; PRT: 60 AA.
AC P52727;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE Metallothionein A (Mt A).
GN MTA.
OS Sparus aurata (Gilthead sea bream).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Perciformes; Percoidae;
OC Sparidae; Sparus.
OX NCBITaxID=8175;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kille P., Olsson P.E.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Tom M.;
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Cousinou M., Lopez-Barea J., Dorado G.;
RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: METALLOTHIONEIN HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -I- DOMAIN: CLASS I METALLOTHIONEIN CONTAINS 2 METAL-BINDING DOMAINS:
CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOATE BRIDGES TO 11
CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
CC -I- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X97276; CA65931.1; -.
CC EMBL: U58774; AAC32738.1; -.
CC EMBL: U93206; AAB51590.1; -.
CC HSSP: P02802; IDFS.
CC InterPro: IPR003019; Metallothion.
CC InterPro: IPR000006; Metallothion_vert.
CC Pfam: PF00131; metalthio.1.
CC PRINTS: PR00860; MTEVERBRATE.
CC PROSITE: PS00203; METALLOTHIONEIN_VRT; 1.
CC Metal-binding; Metal-thiolate cluster; Chelation.
FT DOMAIN 1 28 ALPHA.
FT METAL 4 4 CLUSTER B.
FT METAL 6 6 CLUSTER B.
FT METAL 12 12 CLUSTER B.
FT METAL 14 14 CLUSTER B.
FT METAL 18 18 CLUSTER B.
FT METAL 20 20 CLUSTER B.
FT METAL 23 23 CLUSTER B.
FT METAL 25 25 CLUSTER B.

```

```

FT METAL 28 28 CLUSTER B.
FT METAL 32 32 CLUSTER A.
FT METAL 33 33 CLUSTER A.
FT METAL 35 35 CLUSTER A.
FT METAL 36 36 CLUSTER A.
FT METAL 40 40 CLUSTER A.
FT METAL 43 43 CLUSTER A.
FT METAL 47 47 CLUSTER A.
FT METAL 49 49 CLUSTER A.
FT METAL 54 54 CLUSTER A.
FT METAL 58 58 CLUSTER A.
FT METAL 59 59 CLUSTER A.
SQ SEQUENCE 60 AA: 5966 MW: 6A66F79D02EC591B CRC64;

Query Match 39.3%; Score 51.5; DB 1; Length 60;
Best Local Similarity 47.6%; Pred. No. 1.3;
Matches 10; Conservative 1; Mismatches 7; Indels 3; Gaps 2;

OY 2 CTGAXDCTCTXACTGCGXCP 22
DB 20 CTN-CSCSKCKKSCSC--CP 37

RESULT 14
MT2A_LYCES STANDARD; PRT; 72 AA.
ID MT2A_LYCES
AC 040157;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Metallothionein-like protein type 2 A.
GN MTA.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, ATLSA CRAIG;
RA Whitelaw C.A., Lehuquet J.A., Thurman D.A., Tomsett A.B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L77963; AAB04674.1; -
DR HSSP; P10969; IMGT.
DR InterPro; IPR000347; Metallothion_15.
DR Pfam; PF01439; Metallothio.2; 1.
DR ProDom; PD001611; Metallothion_15; 1.
KW Metal-binding; Metal-thiolate cluster; Chelation; Multigene family.
SQ SEQUENCE 72 AA: 7113 MW: 75FF89BB93AB1893 CRC64;

Query Match 39.3%; Score 51.5; DB 1; Length 72;
Best Local Similarity 40.0%; Pred. No. 1.4;
Matches 10; Conservative 1; Mismatches 11; Indels 3; Gaps 1;

OY 2 CTGAXDCTCTXACTGCGXCPN 23
DB 4 CGGSCGCGSGCKGCGGCGGCMYPD 28

RESULT 15
MT2I_LYCES

```

```

ID MT2I_LYCES STANDARD; PRT; 73 AA.
AC 043513;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Metallothionein-like protein type 2.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, BONNER BESTE; TISSUE=Root;
RA Grilich A., Herdik A., Balzer H., Stephan U., Baumann H.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z68310; CAA92652.1; -
DR InterPro; IPR000347; Metallothion_15.
DR Pfam; PF01439; Metallothio.2; 1.
DR ProDom; PD001611; Metallothion_15; 1.
KW Metal-binding; Metal-thiolate cluster; Chelation; Multigene family.
SQ SEQUENCE 73 AA: 7129 MW: 01BAEB12B25FB9E1 CRC64;

Query Match 39.3%; Score 51.5; DB 1; Length 73;
Best Local Similarity 40.0%; Pred. No. 1.5;
Matches 10; Conservative 1; Mismatches 11; Indels 3; Gaps 1;

OY 2 CTGAXDCTCTXACTGCGXCPN 23
DB 4 CGGSCGCGSGCKGCGGCGGCMYPD 28

```

Search completed: October 21, 2002, 16:38:45
 Job time : 4.82353 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:44 ; Search time 7.76471 Seconds
(without alignments)
534.712 Million cell updates/sec

Title: US-10-032-658-4

Perfect score: 131

Sequence: 1 XCTGXADCTCTACTGCGXCPNA 24

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 126 | 96.2 | 124 | 5 | 09U744 | 09U744 tenebrio mo |
| 2 | 126 | 96.2 | 124 | 5 | 016121 | 016121 tenebrio mo |
| 3 | 126 | 96.2 | 148 | 5 | 016122 | 016122 tenebrio mo |
| 4 | 125 | 95.4 | 112 | 5 | 09U748 | 09U748 tenebrio mo |
| 5 | 124 | 94.7 | 112 | 5 | 09U746 | 09U746 tenebrio mo |
| 6 | 124 | 94.7 | 112 | 5 | 016119 | 016119 tenebrio mo |
| 7 | 124 | 94.7 | 112 | 5 | 016120 | 016120 tenebrio mo |
| 8 | 122 | 93.1 | 112 | 5 | 09U745 | 09U745 tenebrio mo |
| 9 | 110 | 84.0 | 112 | 5 | 09U745 | 09U745 tenebrio mo |
| 10 | 103 | 78.6 | 123 | 5 | 09NC09 | 09nc09 dendroides |
| 11 | 102 | 77.9 | 129 | 5 | 09NC09 | 09nc09 dendroides |
| 12 | 102 | 77.9 | 121 | 5 | 09NC08 | 09nc08 dendroides |
| 13 | 101 | 77.1 | 121 | 5 | 09NC08 | 09nc08 dendroides |
| 14 | 100 | 76.3 | 91 | 5 | 09NC03 | 09nc03 dendroides |
| 15 | 98 | 74.8 | 104 | 5 | 09NC05 | 09nc05 dendroides |
| 16 | 97 | 74.0 | 96 | 5 | 096419 | 096419 dendroides |

| | | | | | | |
|----|------|------|------|----|--------|---------------------|
| 17 | 97 | 74.0 | 98 | 5 | 09NC06 | 09nc06 dendroides |
| 18 | 97 | 74.0 | 108 | 5 | 046351 | 046351 dendroides |
| 19 | 97 | 74.0 | 108 | 5 | 046352 | 046352 dendroides |
| 20 | 97 | 74.0 | 109 | 5 | 046346 | 046346 dendroides |
| 21 | 94 | 71.8 | 136 | 5 | 09NC02 | 09nc02 dendroides |
| 22 | 91 | 69.5 | 118 | 5 | 09NC07 | 09nc07 dendroides |
| 23 | 91 | 69.5 | 122 | 5 | 09NC04 | 09nc04 dendroides |
| 24 | 63 | 48.1 | 1141 | 3 | 074998 | 074998 phaeosphaer |
| 25 | 61 | 46.6 | 99 | 5 | 09N329 | 09n329 caenorhabdi |
| 26 | 59 | 45.0 | 181 | 2 | 09L2D2 | 09l2d2 streptomyce |
| 27 | 59 | 45.0 | 1044 | 3 | 09URM3 | 09urm3 pichia angu |
| 28 | 59 | 45.0 | 1044 | 3 | 000943 | 000943 pichia angu |
| 29 | 58 | 44.3 | 189 | 11 | 09P527 | 09p527 mus musculu |
| 30 | 58 | 44.3 | 196 | 11 | 09D226 | 09d226 mus musculu |
| 31 | 58 | 44.3 | 719 | 4 | 060283 | 060283 homo sapien |
| 32 | 56.5 | 43.1 | 2135 | 4 | 061077 | 061077 plasmodium |
| 33 | 55.5 | 42.4 | 79 | 10 | 09LUX7 | 09lux7 pyrus pyrif |
| 34 | 55.5 | 42.4 | 134 | 5 | 024988 | 024988 giardia lam |
| 35 | 55.5 | 42.4 | 234 | 5 | 09U064 | 09u064 giardia lam |
| 36 | 55.5 | 42.4 | 704 | 5 | 09U048 | 09u048 giardia lam |
| 37 | 55.5 | 42.4 | 709 | 5 | 09XUJ7 | 09xuj7 giardia lam |
| 38 | 54.5 | 41.6 | 66 | 2 | 076959 | 076959 escherichia |
| 39 | 54 | 41.2 | 105 | 4 | 09BYP8 | 09byp8 homo sapien |
| 40 | 54 | 41.2 | 807 | 16 | 09KF73 | 09kf73 bacillus ha |
| 41 | 53.5 | 40.8 | 230 | 11 | 064507 | 064507 mus musculu |
| 42 | 53.5 | 40.8 | 350 | 5 | 094589 | 094589 leishmania b |
| 43 | 53 | 40.5 | 390 | 17 | 027680 | 027680 methanother |
| 44 | 53 | 40.5 | 1110 | 3 | 092198 | 092198 aspergillus |
| 45 | 53 | 40.5 | 1232 | 5 | 096088 | 096088 drosophila |

ALIGNMENTS

RESULT 1
ID 09U744 PRELIMINARY; PRT; 124 AA.

AC 09U744;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM 5-15 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cuculiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF159118; AAD55260.1; -.
DR InterPro: IPR003460; AFP.
DR Pfam; PF02420; AFP; 9.
KW Signal.
FT SIGNAL.
FT CHAIN 29 124
SQ SEQUENCE 124 AA; 13005 MW; 8EC6871047FE2890B CRC64;
POTENTIAL.
THERMAL HYSTERESIS PROTEIN ISOFORM 5-15.

Query Match 96.2%; Score 126; DB 5; Length 124;
Best Local Similarity 87.0%; Pred. No. 2.1e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 CTGXADCTCTACTGCGXCPNA 24
DB 30 CTGXADCTCTACTGCGXCPNA 52

RESULT 2

016121
ID 016121 PRELIMINARY; PRT; 124 AA.
AC 016121;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-3 (3-8) PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF160496; AAB70752.1; -;
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 9.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 124 THERMAL HYSTERESIS PROTEIN ISOFORM YL-3
SQ SEQUENCE 124 AA; 12993 MW; 0DEP87105613791A CRC64;
Query Match 96.2%; Score 126; DB 5; Length 124;
Best Local Similarity 87.0%; Pred. No. 2, 1e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGACDCTCTACTGCGXCPNA 24
DB 30 CTGACDCTCTACTGCGXCPNA 52
RESULT 3
016122
ID 016122 PRELIMINARY; PRT; 148 AA.
AC 016122;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-4 (2-20) PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF160497; AAB70753.1; -;
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 11.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 148 THERMAL HYSTERESIS PROTEIN ISOFORM YL-4
SQ SEQUENCE 148 AA; 15311 MW; EF07B36B8684365 CRC64;
Query Match 96.2%; Score 126; DB 5; Length 148;
Best Local Similarity 87.0%; Pred. No. 2, 5e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGACDCTCTACTGCGXCPNA 24
DB 30 CTGACDCTCTACTGCGXCPNA 52
RESULT 4
090748
ID 090748 PRELIMINARY; PRT; 112 AA.
AC 090748;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM 4-9 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF159114; AAD55256.1; -;
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 4-9.
SQ SEQUENCE 112 AA; 11532 MW; 8A6124C6D31D0E19 CRC64;
Query Match 95.4%; Score 125; DB 5; Length 112;
Best Local Similarity 87.0%; Pred. No. 2, 7e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGACDCTCTACTGCGXCPNA 24
DB 30 CTGACDCTCTACTGCGXCPNA 52
RESULT 5
090746
ID 090746 PRELIMINARY; PRT; 112 AA.
AC 090746;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM 3-4 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;

DR EMBL; AF160497; AAB70753.1; -;
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 11.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 148 THERMAL HYSTERESIS PROTEIN ISOFORM YL-4
SQ SEQUENCE 148 AA; 15311 MW; EF07B36B8684365 CRC64;
Query Match 96.2%; Score 126; DB 5; Length 148;
Best Local Similarity 87.0%; Pred. No. 2, 5e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGACDCTCTACTGCGXCPNA 24
DB 30 CTGACDCTCTACTGCGXCPNA 52
RESULT 4
090748
ID 090748 PRELIMINARY; PRT; 112 AA.
AC 090748;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM 4-9 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF159114; AAD55256.1; -;
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 4-9.
SQ SEQUENCE 112 AA; 11532 MW; 8A6124C6D31D0E19 CRC64;
Query Match 95.4%; Score 125; DB 5; Length 112;
Best Local Similarity 87.0%; Pred. No. 2, 7e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGACDCTCTACTGCGXCPNA 24
DB 30 CTGACDCTCTACTGCGXCPNA 52
RESULT 5
090746
ID 090746 PRELIMINARY; PRT; 112 AA.
AC 090746;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM 3-4 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;

RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL: AF159116; AAD55258.1; -.
DR InterPro: IPR003460; AFP; 8.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 3-4.
SQ SEQUENCE 112 AA; 11637 MW; 58994CCDCFF0F805 CRC64;
Query Match 94.7%; Score 124; DB 5; Length 112;
Best Local Similarity 87.0%; Pred. No. 3.7e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGAGDCTGCTGACTGCGXCPNA 24
DB 30 CTGAGDCTGCTGACTGCGXCPNA 52
RESULT 6
O16119 PRELIMINARY; PRT; 112 AA.
AC O16119;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-1 (2-14) PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL: AF160494; AAB70750.1; -.
DR InterPro: IPR003460; AFP; 8.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM YL-1
SQ SEQUENCE 112 AA; 11558 MW; C06D7AB6401BC909 CRC64;
Query Match 94.7%; Score 124; DB 5; Length 112;
Best Local Similarity 87.0%; Pred. No. 3.7e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGAGDCTGCTGACTGCGXCPNA 24
DB 30 CTGAGDCTGCTGACTGCGXCPNA 52
RESULT 7
O16120 PRELIMINARY; PRT; 112 AA.
AC O16120;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-2 (1-3) PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL: AF160495; AAB70751.1; -.
DR InterPro: IPR003460; AFP; 8.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM YL-2
SQ SEQUENCE 112 AA; 11667 MW; 5883E6DCDF0F805 CRC64;
Query Match 94.7%; Score 124; DB 5; Length 112;
Best Local Similarity 87.0%; Pred. No. 3.7e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGAGDCTGCTGACTGCGXCPNA 24
DB 30 CTGAGDCTGCTGACTGCGXCPNA 52
RESULT 8
O90747 PRELIMINARY; PRT; 112 AA.
AC O90747;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM D-16 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL: AF159115; AAD55257.1; -.
DR InterPro: IPR003460; AFP; 8.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM D-16.
SQ SEQUENCE 112 AA; 11591 MW; BA290B5D8B806C67 CRC64;
Query Match 93.1%; Score 122; DB 5; Length 112;
Best Local Similarity 82.6%; Pred. No. 6.9e-10;
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXCPNA 24
||| :||| ||| ||| |||
DB 30 CTGSDCTCTACTGCGNCNCPNA 52

RESULT 9

ID 090745 PRELIMINARY; PRT; 112 AA.
AC 090745;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM C-9 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm)
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FatBody;
RX MDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF159117; AAD5259.1; -;
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
KW Signal.
FT CHAIN 29 112 POTENTIAL.
SQ SEQUENCE 112 AA; 12002 MW; A4CA9DFRBOAF81FB CRC64;

Query Match 84.0%; Score 110; DB 5; Length 112;
Best Local Similarity 78.3%; Pred. No. 3e-08;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXCPNA 24
||| :||| ||| ||| |||
DB 30 CTGADCTCTACTGCGNCNCPNA 52

RESULT 10

ID 09NC09 PRELIMINARY; PRT; 123 AA.
AC 09NC09;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANTIFREEZE PROTEIN 10 (FRAGMENT).
GN AFP-10.
OS Dendroides canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroides.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;
RT "Isolation and characterization of cDNA clones encoding antifreeze
RT proteins of the pyrochroid beetle Dendroides canadensis.";
RL J. Insect Physiol. 46:365-372(2000).
DR EMBL; AF179414; AAF86365.1; -;
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
FT NON_TER 1 1
SQ SEQUENCE 123 AA; 12882 MW; ASD92CEAD81B4DA5 CRC64;

Query Match 78.6%; Score 103; DB 5; Length 123;
Best Local Similarity 69.6%; Pred. No. 3e-07;
Matches 16; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXCPNA 24
||| :||| ||| ||| |||
DB 21 CTGSDCTCTACTGCGNCNCPNA 43

RESULT 11

ID 09NCRO PRELIMINARY; PRT; 129 AA.
AC 09NCRO;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANTIFREEZE PROTEIN 9.
GN AFP-9.
OS Dendroides canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroides.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;
RT "Isolation and characterization of cDNA clones encoding antifreeze
RT proteins of the pyrochroid beetle Dendroides canadensis.";
RL J. Insect Physiol. 46:365-372(2000).
DR EMBL; AF179413; AAF86364.1; -;
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
SQ SEQUENCE 129 AA; 13679 MW; 5BE91630D6BEF131 CRC64;

Query Match 77.9%; Score 102; DB 5; Length 129;
Best Local Similarity 69.6%; Pred. No. 4.2e-07;
Matches 16; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXCPNA 24
||| :||| ||| ||| |||
DB 27 CTGSDCTCTACTGCGNCNCPNA 49

RESULT 12

ID 09NC08 PRELIMINARY; PRT; 148 AA.
AC 09NC08;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANTIFREEZE PROTEIN 11.
GN AFP-11.
OS Dendroides canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroides.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;
RT "Isolation and characterization of cDNA clones encoding antifreeze
RT proteins of the pyrochroid beetle Dendroides canadensis.";
RL J. Insect Physiol. 46:365-372(2000).
DR EMBL; AF179415; AAF86366.1; -;
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 10.
SQ SEQUENCE 148 AA; 15290 MW; 3A01DED4D282A2CA CRC64;

Query Match 77.9%; Score 102; DB 5; Length 148;
Best Local Similarity 69.6%; Pred. No. 4.7e-07;
Matches 16; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 CTGXADCTCTXACTGCGXCPNA 24
||| :||| ||| ||| |||
DB 27 CTGSDCTCTACTGCGNCNCPNA 49

| | | | | |
|-----------------------|-------|---------------------------|-------|---------------|
| Query Match | 77.1% | Score 101; | DB 5; | Length 121; |
| Best local Similarity | 69.6% | Pred. No. 5, 5e-07; | | |
| Matches | 16; | Conservative | 2; | Mismatches 5; |
| | | | | Indels 0; |
| | | | | Gaps 0; |
| OY | 2 | CTGACDCTCTACGCGXCPCNA | 24 | |
| | | : | | |
| Db | 27 | CTGSDDCSCCTGACCTTCRMCPCNA | 49 | |

| | | | | |
|--------|--|-------------------------|--------------------|------------------|
| | RESULT | 14 | | |
| 09NCR3 | ID | 09NCR3 | PRELIMINARY; | PRT; 91 AA. |
| AC | 09NCR3; | | | |
| DT | 01-OCT-2000 (TREMBLrel, 15, | Created) | | |
| DT | 01-OCT-2000 (TREMBLrel, 15, | Last sequence update) | | |
| DT | 01-JUN-2001 (TREMBLrel, 17, | Last annotation update) | | |
| DE | ANTI-FREEZE PROTEIN 6 (FRAGMENT). | | | |
| GN | AFP-6. | | | |
| OS | Dendroidea candensis. | | | |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | | |
| OC | Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; | | | |
| OC | Cuculiiformia; Pyrochroidae; Dendroidea. | | | |
| OX | NCBI_TaxID=55100; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | Andorfer C.A., Dunan J.G.; | | | |
| RT | "Isolation and characterization of cDNA clones encoding antifreeze | | | |
| RT | proteins of the pyrochroid beetle dendroidea candensis."; | | | |
| RL | J. Insect Physiol. 46:365-372(2000). | | | |
| DR | EMBL: AF179410; AAF86361.1; " | | | |
| DR | InterPro: IPR003460; AFP. | | | |
| DR | Pfam: PF02420; AFP; 6. | | | |
| FT | NON_TER 1 | | | |
| SO | SEQUENCE 91 AA; 9536 MW; 5B9IDEF5283FB16 CRC64; | | | |
| | Query Match | 76.3%; | Score 100; | DB 5; Length 91; |
| | Best Local Similarity | 65.2%; | Pred. NO. 6,1e-07; | |
| | Matches 15; Conservative | 3; | Mismatches 5; | Gaps 0; |
| OY | 2 CTGAACTCCTXACTCGCGCPNA 24 | | | |
| | :: :: :: :: | | | |
| DB | 22 CTGGSDCSCTVSCNTQNCPCNA 44 | | | |

[illegible]

Search completed: October 21, 2002, 16:39:36
Job time : 9.76471 secs

